

مركز البحوث والدراسات

ذكاء الأعمال، والتحليلات، وعلم البيانات

منظور إداري

تأليف

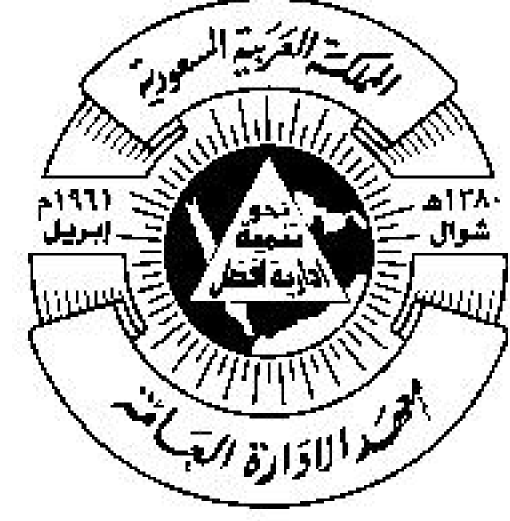
- راميش شاردا
- دورسون ديلن
- إفرايم توربان

ترجمة

د. محمد بن عايض القرني

راجع الترجمة

د. حمود بن ظافر الدوسري



ذكاء الأعمال، والتحليلات، وعلم البيانات

منظور إداري

تأليف

راميش شاردا

دورسون ديلن

إفرايم توربان

ترجمة

د. محمد بن عايض القرني

راجع الترجمة

د. حمود بن ظافر الدوسري

١٤٤١هـ - ٢٠٢٠م

بطاقة الفهرسة

ح معهد الإدارة العامة، ١٤٤١هـ
فهرسة مكتبة الملك فهد الوطنية أثناء النشر.

شاردا ، راميش

ذكاء الأعمال والتحليلات وعلم البيانات: منظور إداري. / راميش شـاردا ؛ دورسـون ديلن ؛ إفرايم توربان ؛ محمد بن عايض القرني. - الرياض، ١٤٤١هـ

٧٦٨ ص ؛ ١٧ × ٢٤ سم.

ردمك: ٩٧٨-٦٠٣-٨٢٧٦-٢٣-٥

١- نظـم المعلومـات الإداريـة ٢- إدارة الأعـمال - معالجة البيانات أ. ديلن، دورسون (مؤلف مشارك) ب. توربـان، إفرايـم (مؤلف مشـارك)، ج. القرني، محمد بن عايض (مترجم) د. العنوان

ديوي: ٦٥٨,٠٢٨٥ ١٤٤١/٩٣١٩

رقم الإيداع: ١٤٤١/٩٣١٩
ردمـــك: ٩٧٨-٦٠٣-٨٢٧٦-٢٣-٥

هذه ترجمة لكتاب:

Business Intelligence, Analytics, and Data Science:

A Managerial Perspective

Ramesh Sharda

Dursun Delen

And Efraim Turban

ملخص المحتويات

المحتويات

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

١٦

موجز تمهيدي:

لقـد أصبـح تحليل البيانات هـو المنهج الرائد في تقنيـة العقد الحالي. وتقـوم شركات مثل أي بي إم (IBM)، وإس إيـه بي (SAP)، وإس إيـه إس (SAS)، وتيراداتـا (Teradata)، وأوراكل (Oracle)، ومايكروسـوفت (Microsoft)، وديل (Dell)، وغيرها من الشركات بإنشـاء وحدات تنظيمية جديدة تركِّز على التحليلات التي تسـاعد الشركات على أن تصبح أكثر فعاليةً وكفاءةً في عملياتها. ويستخدم صانعو القرار مزيدًا من الأدوات المحوسبة لدعم أعمالهم؛ بل إنَّ المستهلكين يسـتخدمون أدوات تحليلية، بشـكلٍ مباشر أو غير مباشر؛ من أجل اتخاذ قرارات بشأن الأنشطة المعتادة كالتسـوق والصحة/ الرعاية الصحية والسـفر والترفيه. وقد تطـوَّر مجالا ذكاء الأعمال BI وتحليـلات الأعمال BA بسرعة هائلة ليصبحا أكثر تركيزًا على التطبيقات المبتكرة لاسـتخلاص المعلومات من تيارات البيانات التي لم يكن من الممكن الحصولُ عليها في وقت ما في الماضي أو كان تحليلها على نحوٍ أقل بكثير حتى باستخدام أيٍّ وسيلة متطوِّرة في هذا الوقت. وتظهر التطبيقات الجديـدة يوميًّا في مجال الرعاية الصحية، والرياضة، والسـفر، والترفيه، وإدارة سلاسـل التوريد، والمرافق، وكل صناعة تقريبًا قد تتخيلها. ومن ثَمَّ أصبح مصطلح التحليلات سائدًا وتطوَّر بالفعل إلى مصطلحات أخرى مثل علم البيانات وهو أحدث تجسيدٍ للتعلُّم العميق وإنترنت الأشياء.

وتُقـدِّم هذه الطبعة منظـورًا إدارياً لتحليل الأعمال، بدءًا مـن التحليلات الوصفية (مثل: طبيعـة البيانات، والنمذجة الإحصائية، وتصوير البيانات، وذكاء الأعمال) لينتقل إلى التحليلات التنبؤية (مثل: التنقيب في البيانات، والتنقيب في النص/ الويب، والتنقيب في وسـائل التواصل الاجتماعـي)، ومن ثَمَّ إلى التحليلات التوجيهية (مثل: التحسـين والمحـاكاة)، وانتهاءً بالبيانات الضخمة، والاتجاهات المسـتقبلية، والخصوصية، والاعتبـارات الإدارية. ويدعم الكتابُ موقع ويب (pearsonhighered.com/sharda). وكذلك الموقع المسـتقل dssbibook. com. كما سنوفر روابط إلى برامج تعليمية من خلال قسم خاص لمواقع الويب.

إن الغـرض من هـذا الكتاب؛ هو تزويد القارئ بهذه التقنيات التي تُسـمَّى عمومًا تحليلات الأعـمال أو علـم البيانات؛ ولكنها كانت معروفةً من قبل بأسـماء أخـرى. ويعرض هذا الكتاب أساسـيات الأسـاليب والطريقة التي يتمُّ من خلالها إنشـاء هذه الأنظمة واستخدامها. ونتبع في ذلـك نهج EEE لتقديم هـذه المواضيع: Exposure (التعـرض)، وExperience (التجريب)، و Exploration (الاستكشـاف). ويوفر الكتابُ في المقام الأول التعرُّض لمختلف أساليب التحليلات وتطبيقاتها. والفكرة هي أن الطالب سـوف يكون مصدر إلهام لتعلُّم كيفية اسـتخدام المنظمات

الأخـــرى للتحليـــلات؛ من أجل اتخاذ القـــرارات أو للحصول على ميزة تنافســـية. ونعتقد أن مثل هـــذا التعرض لما يحدث باســـتخدام التحليلات وكيفية تحقيقه هو المكون الرئيس للتعلم بشـــأن التحليلات. وفي وَصْف الأســـاليب، نعرض كذلك أدوات البرمجيات المحدَّدة التي يمكن استخدامها لتطوير مثل هذه التطبيقات. ولا يقتصر الكتاب على أداة برمجية واحدة، حتى يستطيع الطالب تجريب هذه الأســـاليب باســـتخدام أيِّ عددٍ من أدوات البرمجيـــات المتاحة. كما نقدِّم اقتراحات مُحدَّدة في كلِّ فصل حتى يتســـنى للطالبِ والأســـتاذ اســـتخدامُ هذا الكتـــاب الذي يحتوي على العديد من أدوات البرمجيات المختلفة. وسيتضمَّن موقع الويب المصاحب لكتابنا أدلة لبرمجيات محددة، ولكن يمكن للطلاب اكتســـابُ الخبرة في هذه الأســـاليب بطرق عديدة ومختلفة. وأخيراً؛ نأمل أن يُمكِّن كلٌّ من التعرض والتجريب، ويحفزا القراء على استكشـــاف إمكانات هذه الأساليب كل في مجاله. ولتســـهيل مثل هذا الاستكشـــاف؛ نقوم بتضمين التجارب التي توجههم إلى شـــبكة جامعـــة تيراداتا Teradata وغيرها من المواقع، والتي تشـــمل تدريبات قائمـــة على عمل الفريق إذا لزم الأمر. وسنُســـلِّط الضوء أيضاً على التطبيقات الجديدة والمبتكرة التي نتعلمها على موقع الكتاب على شبكة الإنترنت.

وتركز معظم التحســـينات المحدَّدة التي أُدخِلَت في هـــذه الطبعة الرابعة على أربعة مجالات هي: إعادة التنظيم، والفصول الجديدة، وتحديث المحتوى، وتركيز أكثر وضوحاً. وعلى الرغم من التغييرات العديدة؛ فقد حافظنا على الشمولية وسهولة الاستخدام التي جعلت الكتاب من الكتب الرائدة في السوق. وأخيرًا؛ نقدم موادًا دقيقةً ومحدثةً غير متوفرة في أيِّ نص آخر.

ما الجديدُ في الطبعة الرابعة؟

ولأننـــا نهدف إلى تحســـين النـــص؛ قمنا في هذه الطبعـــة بإعادة تنظيم جوهريـــة للنص نركِّز مـــن خلالها عـــلى تحليلات الأعمال. وقد تمَّ تنظيمُ هذه الطبعة لتتضمن ثلاثة أنواع رئيســـة من تحليلات الأعمال (وهي: الوصفية، والتنبؤية، والتوجيهية). وتحتوي الطبعة الجديدة على العديد مـــن الإضافات التي تتناســـب مع التطورات الحديثة، وقد تمَّ حـــذفُ المحتويات التي أصبحت لا تتلاءم مع هذه التطوُّرات. وقد أُجرِيَت التغييراتُ الرئيسية التالية:

- **التنظيم الجديد:** يقرُّ هذا الكتابُ بثلاثةِ أنواعٍ من التحليلات: الوصفية، والتنبؤية، والتوجيهية؛ وهـــو التصنيف الذي يروِّج له معهدُ بحوث العمليـــات وإدارة العلوم إنفورمز (INFORMS). يقدِّم الفصل الأول ذكاء الأعمال والتحليل مع التركيز على تطبيقاتهما في العديد من الصناعات. كما يتضمَّن هذا الفصلُ كذلك لمحةً عامةً عن النظام البيئي للتحليل لمســـاعدة المســـتخدم على

استكشـاف جميع الطرق المختلفة التي يمكن المشـاركة فيها وتنميتها في بيئة التحليلات. ويلي ذلك نظـرةٌ عامة على الإحصاءات وأهميـة البيانات والتحليـلات الوصفية/ التصوير في الفصل الثاني. ويتناول الفصل الثالث مسـتودع البيانات وأُسُـس البيانات بما في ذلك المحتوى المُحدَّث؛ وخصوصًـا بحيرات البيانات. ويغطـي الفصل الرابع التحليلات التنبؤيـة. أمَّا الفصل الخامس فيوسِّـع من نطاق تطبيق التحليلات إلى النصِّ والمواقع الإلكترونية ووسائل التواصل الاجتماعي. ويتناول الفصل السادس التحليلات التوجيهية؛ وخاصةً البرمجة الخطية والمحاكاة. وهو محتوى جديـد تمامًا بهذا الكتاب. ويقدِّم الفصل السـابع أدوات ومِنصَّـات البيانات الضخمة. ويُختتَم الكتـابُ بالفصل الثامن والاتجاهـات والمواضيع الحديثة الظهور في تحليلات الأعمال بما في ذلك تحليلات الموقع، وإنترنت الأشـياء والاعتبـارات الخصوصية/ الأخلاقيـة في التحليلات. وتعترف مناقشة النظام البيئي التحليلي بالتحليلات التوجيهية كذلك.

- **الفصول الجديدة:** تمَّت إضافة الفصول التالية:
  - الفصـل الثاني: التحليلات الوصفية (١): طبيعـة البيانات، النمذجة الإحصائية، وتصوير البيانات ويهدف هذا الفصلُ إلى وضع حجر الأسـاس للدراسة التحليلية عن طريق فهم شـامل لطبيعة البيانات، وهو العنصرُ الرئيس لأيِّ دراسـة تحليليـة. ويلي ذلك، عرض النمذجـة الإحصائية كجزءٍ من التحليلات الوصفية. ولقـد أصبح تصويرُ البيانات جزءًا شائعًا من أيِّ تقريرٍ عن الأعمال و / أو أيِّ مشروع تحليلي وصفي؛ وبالتالي، نقوم بشرح ذلـك بالتفصيـل في هذا الفصل. ونعـزِّز الفصل بالعديد من الحـالات والأمثلة في عالم الواقع (٧٥٪ من المواد الجديدة).
  - الفصل السـادس: التحليلات التوجيهية: التحسـين والمحاكاة، ويقدِّم هذا الفصل مادةً تحليليةً وصفيةً لهذا الكتاب. ويركِّز الفصل على نمذجة التحسـين في إكسيل باستخدام تقنيات البرمجة الخطية. كما يقدِّم مفهوم المحاكاة. والفصل هو نسـخةٌ محدثةٌ لمواد فصلين في كتابنا DSS، الطبعة العاشرة. وهو بالنسـبة لهـذا الكتاب فصلٌ جديدٌ تمامًا (٩٩٪ مواد جديدة).
  - الفصل الثامن: الاتجاهات المسـتقبلية، والخصوصية، والاعتبارات الإدارية في التحليلات ويتنـاول هذا الفصـلُ العديد من الظواهر الجديدة التي تغـيِّر أو من المحتمل أن تغير التحليـلات. حيث يتنـاول تحليلات الجغرافيا المكانية، وإنترنت الأشـياء، وتحديثًا كبيرًا للمـواد المتعلقة بالتحليلات المسـتندة إلى الحوسـبة السـحابية. كما يقـوم بتحديث ما تناولته الطبعة الأخيرة بشأن الاعتبارات الأخلاقية والعملية (٧٠٪ مواد جديدة).

- **الفصــول المنقحــة:** وقد تمَّ تنقيحُ جميع الفصول الأخــرى وتحديثها كذلك. وفيما يلي ملخصٌ للتغيرات في هذه الفصول الأخرى:

  - الفصل الأول: نظرة عامة على ذكاء الأعمال، والتحليلات، وعلم البيانات: تمَّت إعادة كتابة هذا الفصل وتوسعته بشــكلٍ كبير؛ بحيث يُفتتح بمقالة قصيرة جديدة تغطي تطبيقات متعــددة من التحليلات في مجال الرياضة. وهو يقدِّم الأنواع الثلاثة من التحليلات على النحو الذي اقترحته مؤسســة إنفورمز (INFORMS): التحليــلات الوصفية، التنبؤية، والتوجيهية. وكما أشرنا ســابقاً؛ فإن هذا التصنيف يُســتخدَم في التوجيه لإعادة تنظيم كاملة للكتاب نفســه (المحتوى الســابق؛ ولكن بشــكلٍ جديد). ثم يتضمن عِدة أمثلة جديــدة من التحليلات في مجــال الرعاية الصحية وفي قطاع تجارة التجزئة. وأخيراً؛ فإنه يُختتم بتغطية مُوسَّعة ومُحدَّثة بشكلٍ كبير للنظام البيئي للتحليلات لمنح الطلاب شعورًا باتساع نطاق صناعة التحليلات وعلم البيانات (حوالي ٦٠٪ مواد جديدة).

  - الفصل الثالث: التحليلات الوصفية (٢): ذكاء الأعمال ومســتودع البيانات: ويُعَدُّ هذا الفصل فصلًا قديمًا أُضِيف إليه بعضُ الأقســام الفرعية الجديدة (مثل: بحيرات البيانات) وحالات جديدة (حوالي ٣٠٪ مواد جديدة).

  - الفصــل الرابــع: التحليلات التنبؤيــة (١): عملية، وطرق، وخوارزميــات التنقيب في البيانــات، وهذا أيضاً فصلٌ قديمٌ مع تنظيم/ تدفــق جديد للمحتوى وبعض الحالات الجديدة (حوالي ٢٠٪ مواد جديدة).

  - الفصــل الخامــس: التحليلات التنبؤية (٢): تحليلات النص، والويب، ووســائل التواصل الاجتماعي وهو فصلٌ قديم مع تنظيم/ تدفق جديد للمحتوى وبعض الحالات الجديدة (حوالي ٢٥٪ مواد جديدة).

  - الفصــل الســابع: مفاهيم وأدوات البيانــات الضخمة: كان هــذا الفصل هو الفصل الســادس في الطبعة الســابقة. وقد تمَّ تحديثه بمقدمة افتتاحيــة مختصرة وحالات جديــدة، وتغطية لتيراداتا أســتر (Teradata Aster)، ومواد جديدة تتعلق بالبيانات البديلة (حوالي ٢٥٪ مادة جديدة).

- **فريق تأليف مجدد:** اســتنادًا إلى المحتوى المتميز الذي قام بإعداده مؤلفو الطبعات الســابقة (Turban، وSharda، وDelen، وKing)؛ قام بتنقيح هذه الطبعة كلٌّ من Ramesh Sharda

وDursun Delen. وقـد عمـل الأخيران على نطاق واسـع في التحليلات بمـا لهما من كفاءة في المجال الصناعي فضلاً عن الخبرة البحثية.

- **طباعة ملونة:** إننا سـعداء حقاً أن يظهر هذا الكتاب بالألوان. وحتى الأشـكال التي وردت في النسـخ السابقة أُعِيد رسـمها حتى تتمتع بميزة الألوان. إنَّ اسـتخدام اللون يُعزِّز العديد من الأمثلة التصويرية والمواد الأخرى كذلك.
- **موقع ويب مُحدَّث ونشـط:** يمكن لقراء هذا الكتاب الآن الوصول إلى موقع على شـبكة الإنترنت يشتمل على روابط خاصة بالأخبار، والبرمجيات، والبرامج التعليمية، وحتى مقاطع فيديو يوتيوب ذات الصلة بالمواضيع التي يغطيها الكتاب. وهذا الموقع سوف يكون متاحًا في dssbibook.com.
- **محتوى مُنقَّح ومُحدَّث:** لقد استهللنا جميع الفصول تقريبًا بمقدمات تستند على القصص والأحداث الجارية. وبالإضافة إلى ذلك؛ تمَّ تحديث الحالات العملية في جميع فصول الكتاب لتشمل الأمثلة الحديثة لتطبيقات أسـاليب/ نماذج معينة. وقد تمَّت إضافة روابط موقع ويب جديدة في جميع فصـول الكتـاب. كما حذفنا العديد من الروابط والمراجع القديمـة. وأخيراً؛ فقد أوردنا في معظم الفصول تدريبات ومهام جديدة على الإنترنت، وأسئلة مناقشة في جميع فصول الكتاب.
- **روابط إلى شـبكة جامعة تيراداتا (TUN):** وتتضمن معظم الفصول روابط جديدة إلى شـبكة جامعة تيراداتا (teradatauniversitynetwork.com).
- **عنوان الكتاب:** وكما هو واضحٌ بالفعل؛ فقد تغير عنوان الكتاب وتركيزه بشكلٍ كبير.
- **دعم البرمجيات:** يوفر موقع TUN على شـبكة الإنترنت دعم البرمجيات دون أيِّ رسـوم. كما يوفر روابط إلى التنقيب في البيانات مجانًا وغيرها من البرامج. وبالإضافة إلى ذلك؛ يوفر الموقع تدريبات على استخدام مثل هذه البرامج.
- **الباقـة الإضافية www.pearsonhighered.com/sharda:** هي مجموعة شـاملة ومرنة من الدعـم التقني لتعزيـز تجربة التعليم والتعلم. كـما تتوفَّر المواد الإضافية لـكلٍّ من الطالب والمدرب على موقع الكتاب الإلكتروني، pearsonhighered.com/sharda:

١- **دليـل المدرب:** يتضمَّن دليلُ المدرب أهداف التعلُّم بكل فصل، واقتراحات التدريس (بمـا في ذلك تعليمات للمشـاريع). ويتوفـر دليلُ المدرب في قسـم هيئة التدريس في pearsonhighered.com/sharda.

٢- **ملـف بنـد الاختبار وبرنامـج TestGen:** إن ملف بندِ الاختبـار عبارة عن مجموعة شـاملة من أسـئلة (صح / خطأ)، وأسـئلة الاختيار المتعدد، واملأ الفراغ، وأسئلة مقالية.

ويتم تصنيف الأسـئلة حسب مستوى الصعوبة، ويُشار إلى الإجابات برقم صفحة الكتاب التـي توجد بها. ويتوفَّر ملف بند الاختبـار في Microsoft Word وTestGen. ويتوفر برنامـج توليـد اختبـار Pearson Education مـن /www.pearsonhighered.com irc. ويتوافق البرنامج مع كلٍّ من الحاسـب الشـخصي وأجهـزة MAC، ويتمُّ تنزيله مع جميع أسئلة ملف بند الاختبار. ويمكنك أن تقوم يدوياً أو عشوائياً بعرض أسئلة الاختبار والسحب والإسقاط لإنشـاء اختبار. ويمكنك إضافة أو تعديل أسئلة بنك الاختبار حسب الحاجة. ويتم تحويل برمجية TestGens حتى يمكن استخدامها في كلٍّ من BlackBoard وWebCT وMoodle وD2L وAngel. ويمكـن العثـور عـلى هذه النسـخ على موقع pearsonhighered.com/sharda. كما يتوفر برنامج TestGen في Respond us ويمكن الاطلاع على www.respondus.com.

٣- **شـرائح باوربوينت:** تتوفر شرائح باوربوينت التي تتميز بخاصية الإضاءة؛ بناءً على المفاهيم الأساسية في النص. ويمكن لأعضاء هيئة التدريس تنزيل شرائح باوربوينت من pearsonhighered.com/sharda.

شكر وتقدير:

قـدَّم العديدُ من الأفراد اقتراحات وانتقادات منذ نـشر الطبعة الأولى من هذا الكتاب. وقد شـارك العشراتُ من الطلاب في الاختبارات الصفية لمختلف الفصول والبرامج والمشكلات وساعدوا في جمع المواد. ولا يمكن ذِكر اسم كلٍّ من شارك في هذا المشروع؛ ولكننا نوجِّه الشكر لهم جميعاً. كما قدَّم البعض مساهمات كبيرة، هؤلاء يستحقون تقديراً خاصًّا.

أولاً، نقـدِّر جهود الأفـراد الذين قاموا بمراجعات رسـمية للطبعات مـن الأولى حتى الثالثة (الانتماءات المدرسية اعتباراً من تاريخ المراجعة):

- آن أكسوت، كلية مجتمع بيدمونت المركزية.
- باي أرينز، جامعة دريكسل.
- أندي بورشرز، جامعة ليبسكومب.
- رانجيت بوس، جامعة نيو مكسيكو.
- مارتي كروسلاند، جامعة أمريكا الوسطى الناصرية.
- كورت إنجمان، كلية أيونا.

- بديع فرح، جامعة ميشيغان الشرقية.
- غاري فارار، كلية كولومبيا.
- جيري فيرميستاد، معهد نيو جيرسي للتكنولوجيا.
- كريستي م. فولر، جامعة لويزيانا التقنية.
- مارتن غروسمان، كلية بريدج ووتر الحكومية.
- جهانكير كريمي، جامعة كولورادو، دنفر.
- هوي لي، جامعة ميشيغان الشرقية.
- ناتالي نازارينكو، جامعة نيويورك فريدونيا الحكومية.
- جو إنج لي بارتريدج، جامعة كونيتيكت الحكومية المركزية.
- غريغوري روز، جامعة واشنطن الحكومية، فانكوفر.
- خواجة سعيد، جامعة ويتشيتا الحكومية.
- كالا تشاند سيل، جامعة لويولا ماريماونت.
- جوشوا س. وايت، دكتوراه، جامعة نيويورك الحكومية معهد الفنون التطبيقية.
- روجر ويلسون، جامعة فيرمونت الحكومية.
- فنسنت يو، جامعة ميسوري للعلوم والتكنولوجيا.
- فان تشاو، جامعة ساحل خليج فلوريدا.

ونقدِّر أيضاً جهود الأفراد الذين قاموا بمراجعات رسـمية لهـذا النص وكتابنا الآخر DSS - ذكاء الأعمال والتحليلات: نظم دعم اتخاذ القرار، الطبعة العاشرة، بيرسون للتعليم، ٢٠١٣م.

ثانياً، أسـهم العديدُ من الأفراد بمواد في النص أو المواد الداعمة. وقد قدَّمت سـوزان باسكن مـن تيراداتا والدكتور ديفيد شريدر مسـاعدةً خاصة في تحديد المحتوى الجديد لكلٍّ من TUN وتيراداتـا للكتاب وترتيب الأذونات لنفس الغرض. كما أسـهم الدكتـور ديف شريدر في افتتاح المقالـة القصيرة للكتاب. تضمَّنت هذه المقالة القصيرة أيضاً مواد طوَّرها الدكتور أشـيش جوبتا من جامعة أوبورن وجاري ويلكيرسـون من جامعة تينيسي تشـاتانوجا. وسـوف توفر مقدمة رائعـة للتحليلات. نشـكر أيضـاً INFORMS على إذنهم بتسـليط الضوء عـلى المحتوى من الواجهـات. كما نعترف بدور الأفراد التالية أسـماؤهم في المسـاعدة في تطوير هذا الإصدار من الكتاب؛ وهم: بانكوش كالجوترا، وبراسون ماثور، وروبيش أجاروال، وشوبهام سينغ، ونان ليانغ،

وجاكوب بيرسون، وكينزي كليمير، وإيفان مورليت (وجميعهم من جامعة أوكلاهوما الحكومية). وقـد قام فريق Teradata Aster، وخاصةً مارك أووت بتوفير المـواد اللازمة للمقالة الافتتاحية للفصل السـابع. إذ يتم تكييفُ مواد Aster في الفصل ٧ من أدلة التدريب الأخرى التي طوَّرها جـون ثوما وجريـج بيثاردي. كما قاد الدكتـور برايان ليكلاير، مدير قسـم المعلومات في شركة Humana بمسـاهمات من العديد من دراسـات الحالة الصحية الواقعية التي طوَّرها فريقه في Humana. وأسـهم ابهيشـيك راثي من vCreaTek برؤيته للتحليـلات في صناعة بيع التجزئة. كما نتوجه بالشـكر والتقدير للدكتور ريك ويلسـون على التدريبات الممتازة لتدريس وممارسة مهارات البرمجة الخطية في Excel. كما سمح لنا مات تارك بتكييف مواد النظام البيئي لإنترنت الأشياء الخاصة به. ويقرُّ راميش بالمساعدة التي قدَّمتها ابنته، رشي شاردا سن في تحرير النسخ. بالإضافة إلى ذلك؛ قدَّم طلابُ الدكتوراه وزملاؤنا من الباحثين التالي أسماؤهم محتوًى أو نصيحةً ودعمًا للكتاب بعدة طرق مباشرة وغير مباشرة:

- أسيل أوزتيكين، جامعة ماساتشوستس - لويل.
- إينيس إريارسوي، جامعة سيهير.
- حامد مجيدي زولبانين، جامعة بول الحكومية.
- أمير حسن زاده، جامعة رايت الحكومية.
- سوبافيتش (فون) بنجنات، جامعة داكوتا الشمالية الحكومية.
- كريستي فولر، جامعة بويز الحكومية.
- دانييل أساmواه، جامعة رايت الحكومية.
- سليم زايم، جامعة إسطنبول الفنية.
- نيهات كاساب، جامعة سابانسي.

ثالثاً. بالنسبة للنسخة السابقة، نقر بمساهمات ديفيد كينج (JDA Software Group, Inc.). ومن بين المساهمين الأساسيين الآخرين في الإصدار السابق جي أرونسون (جامعة جورجيا)، والذي كان مؤلفًا مشاركًا، والمساهم في فصل مستودعات البيانات؛ ومايك جول (جامعة أريزونا الحكومية)، والتي تمَّ إدراجُ إسهاماتها في الفصل ١؛ وكذلك تي بي ليانج (جامعة صن - يت سِن، تايوان)، والذي أسهم بمواد حول الشبكات العصبية في الإصدارات السابقة. كما تعاونت جودي لانج معنا جميعًا في توفير التحرير، وقادتنا خلال المشروع بأكمله في الإصدار الأول.

رابعـــاً، تعاون العديدُ من الموردين من خلال توفير دراســات الحالــة و / أو برمجيات العرض التوضيحي للإصدارات السابقة وهم: أكسيوم (ليتل روك، أركنساس)، برمجيات كاليفورنيا العلمية (نيفادا ســيتي، كاليفورنيا)، كاري هاروين من تطوير كاتاليست (يوكا فالي، كاليفورنيا)، آي بي إم (سان كارلوس، كاليفورنيا)، شركة دي إس جروب، (جرينيتش، كونيكتيكت)، جريجوري بياتيتسكي شابيرو من KDnuggets.com، جاري لين من NeuroDimension Inc. (جاينيسفيل، فلوريدا)، برمجيات باليســاد (نيوفيلد، نيويورك)، بروميســد لاند للتقنيات (نيو هافن، كونيكتيكت)، نظم سالفورد (لا جولا، كاليفورنيا)، سينس الشبكات (نيويورك، نيويورك)، جاري ماينر من شركة ستات ســوفت، (تولسا، أوكلاهوما)، شركة وارد سيســتمز جروب، إنك (فريدريك، ميريلاند)، شركة نظم آيديا فيشر، (إيرفينج، كاليفورنيا)، ونظم ورد تك (أوريندا، كاليفورنيا).

خامســاً، شكرٌ خاص لشــبكة جامعة تيراداتا وخاصة سوزان باســكن، مديرة البرنامج؛ وهيوج واتســون الذي أسَّس شــبكة جامعة تيراداتا، ومايكل جول، وبارب ويكســوم، وماري جروس على تشجيعهم لنا على ربط هذا الكتاب بشبكة جامعة تيراداتا وتوفير مواد مفيدة له.

أخيراً؛ نثني على فريق بيرســون: سامانتا لويس، التي عملت معنا على هذه المراجعة وقامت بتنظيــم الإخراج الملون للكتاب؛ وفريــق الإنتاج، آن بوليدو، وريفاتي فيســواناثان والموظفون في سينفيو، الذين حوَّلوا المخطوطة اليدوية إلى كتاب.

نودُّ أن نشكر كلَّ هؤلاء الأفراد والشركات. فبدون مساعدتهم، لما كان إنشاء هذا الكتاب ممكنًا.

راميش شاردا

دورسون ديلن

إفرايم توربان

* لاحــظ أنَّ روابــط موقع الويب ديناميكية؛ فبمجرد مثول هذا الكتــاب للطباعة، تمَّ التحقق من أنَّ جميع مواقع الويب المذكورة نشــطة وصالحة. ولكن في بعض الأحيان تتغير أو تقف مواقع الويب المشــار إليها في النص بســبب تغيير الشركات لأســمائها، أو قد يتمُّ شراؤها أو بيعها أو دمجها أو فشــلها، وفي أحيان أخرى تكون مواقع الويب معطلةً للصيانة أو الإصلاح أو إعادة التصميم. وفي حين أنَّ معظم المنظمات قد أســقطت البادئة «www» من مواقعها؛ فإنَّ البعض الآخر لا يزال يســتخدمها. إذا واجهت مشــكلةً في الاتصال بموقع ويب تمَّ الإشــارة إليه، فضلاً تحلَّى بالصبر وقم ببساطة بتشغيل بحث الويب لمحاولة تحديد الموقع الجديد؛ إذ إنه في معظم الأوقات يمكن العثورُ على الموقع الجديد بسرعة. نعتذر مقدماً عن هذا الإزعاج.

ذكاء الأعمال، والتحليلات، وعلم البيانات - منظور إداري:

يتناول هذا الكتاب مجموعة من تقنيات الحاسب التي تدعم العمل الإداري؛ وبشكل أساسي صنع القرار. وقد كان لهذه التقنيات أثر عميق على إستراتيجية الشركة وأدائها وقدرتها التنافسية. وتُسمى هذه التقنيات مجتمعةً ذكاء الأعمال وتحليلات الأعمال وعلم البيانات. وعلى الرغم من مناقشة تطور المصطلحات، يتم استخدام هذه الأسماء أيضاً بالتبادل. ويروي هذا الكتاب قصصًا عــن كيفية توظيف الأفراد الأذكياء لهذه التقنيات لتحســين الأداء والخدمة والعلاقات في الأعمال والحكومات والعوالم غير الربحية.

نُبذة عن المؤلفين:

- **راميش شاردا Ramesh Sharda:** (حاصل على درجة ماجستير إدارة الأعمال، والدكتوراه، من جامعة ويسكونسن - ماديسون)، وهو نائب العميد لبرامج البحث والخريجين، وعضو في واتسون / كونوكوفيليبس، وأستاذ أعضاء مجلس الجامعة للعلوم الإدارية ونظم المعلومات في مدرسة سبيرز للأعمال بجامعة أوكلاهوما الحكومية (OSU). قام بالتأسيس والإشراف على درجة الدكتوراه في إدارة الأعمال في برنامج مديري التنفيذ. وقد تمَّ نشرُ حوالي ٢٠٠ ورقة تصِف أبحاثه في كبرى المجلات، بما في ذلك بحوث العمليات، وعِلم الإدارة، وبحوث نظم المعلومات، ونظم دعم القرار، ومجلة نظم المعلومات الإدارية. كما قام بتأسيس AIS SIG على أنظمة دعم القرار وإدارة المعرفة (SIGDSA). يعمل الدكتور شاردا في العديد من مجالس التحرير، بما في ذلك مجلة علوم القرار، ونظم دعم القرار، وقاعدة بيانات ACM. قام بتأليف وتحرير العديدِ من الكتب النصية والكتب البحثية، كما يشارك في تحرير العديد من كتب سلسلة Springer (وهي سلسلةٌ متكاملةٌ في نظم المعلومات، وواجهات بحوث العمليات/ علوم الحاسب، وسجلات نظم المعلومات). ويشغل حالياً منصب المدير التنفيذي لشبكة جامعة تيراداتا. وتتمثل اهتماماته البحثية الحالية في نظم دعم القرار وتحليلات الأعمال وتقنيات إدارة الحمل الزائد للمعلومات.
- **دورسون ديلن Dursun Delen:** (حاصلٌ على درجة الدكتوراه، من جامعة أوكلاهوما الحكومية) وهو عضو سبيرز إندويد في إدارة الأعمال، وعضو في مؤسسة باترسون في تحليلات الأعمال، ومدير أبحاث مركز تعاون النظم الصحية، وأستاذ أعضاء مجلس الجامعة للعلوم الإدارية ونظم المعلومات في مدرسة سبيرز للأعمال بجامعة أوكلاهوما الحكومية (OSU). وقد عمل قبل مسيرته الأكاديمية عالم أبحاث لمدة ٥ سنوات، في شركة خاصة للأبحاث والاستشارات، وهي شركة Knowledge Based Systems، في College Station بولاية تكساس، وقد قاد خلال تلك الفترة عدداً من نظم دعم القرار ونظم المعلومات الأخرى، ومشاريع البحوث ذات الصِّلة التي تموِّلها عدة وكالات فيدرالية، مثل وزارة الدفاع (DoD)، وإدارة الطيران والفضاء الوطنية (NASA)، والمعهد الوطني للمعايير والتقنية (NIST)، ومنظمة دفاع القذائف البالستية (BMDO)، ووزارة الطاقة (DOE). نشر الدكتور ديلن أكثر من ١٠٠ مقالة استعرضها الزملاء، وقد ظهر بعضها في كبرى

المجلات مثل علوم القرار، وأنظمة دعم اتخاذ القرار، واتصالات ACM، وبحوث الحاسب والعمليات، وأجهزة الحاسب في الصناعة، ومجلة إدارة عمليات الإنتاج، والذكاء الاصطناعي في الطب، والمجلة الدولية للمعلوماتية الطبية، والأنظمة الخبيرة مع التطبيقات، والاتصالات اللاسـلكية IEEE. وقام مؤخراً بتأليف / المشاركة في تأليف سبعة كتب نصية في مجالات واسـعة النطاق من تحليـلات الأعمال، والتنقيب في البيانـات، والتنقيب في النص، وذكاء الأعـمال، وأنظمة دعم القرار. وغالباً ما يُدعى للمشـاركة في المؤتمرات الوطنية والدولية لعناوين رئيسـة عن موضوعات ذات صِلة بالتنقيب في البيانات/ النص وتحليلات الأعمال ونظـم دعـم القرار وذكاء الأعمال وإدارة المعرفة. شـغل منصب مسـاعد الرئيس العام للمؤتمر الدولي الرابع للحوسبة الشبكية وإدارة المعلومات المتقدمة (٢-٤ سبتمبر ٢٠٠٨م، في سيول، كوريا الجنوبية) ويشارك بانتظام في رئاسة وتسيير أعمال مختلف مؤتمرات نظم المعلومات والمؤتمرات التحليلية. يشـغل حاليًّـا منصب رئيس التحرير، محرر أول، محرر مشـارك، أو عضـو في هيئة تحرير لأكثر من عشرة مجـلات أكاديمية. ويعمل حاليًّا رئيس تحرير، أو محرر أول، أو محرر مشارك، أو عضو مجلس تحرير لأكثر من اثنتي عشرة مجلة أكاديمية. وتتمثل اهتماماته البحثية والتعليمية في التنقيب في البيانات والنص، وتحليلات الأعمال، ونظم دعم القرار، وإدارة المعرفة، وذكاء الأعمال، ونمذجة المؤسسات.

- **إفرايم توربان Efraim Turban:** (حاصل على درجة ماجستير إدارة الأعمال، والدكتوراه، مـن جامعة كاليفورنيا، بولايـة بيركلي) وهو باحث زائر في معهد الباسـيفيك لإدارة نظم المعلومات، بجامعة هاواي. وعمل قبل ذلك في العديد من الجامعات، مثل: جامعة سيتي في هونغ كونغ، وجامعة لي هاي، وجامعة فلوريدا الدولية، وجامعة كاليفورنيا الحكومية، ولونج بيتـش، وجامعة إلينوي الشرقية، وجامعة كاليفورنيـا الجنوبية. وقد قام الدكتور توربـان بتأليـف أكثر من ١٠٠ ورقة محكمـة تم نشرها في المجلات الرائـدة، مثل: علوم الإدارة، ونظـم المعلومات الإدارية الفصلية، ودعـم اتخاذ القرار. كما ألَّف ٢٠ كتاباً، مثل التجارة الإلكترونية: منظور إداري وتقنية المعلومات للإدارة. وهو أيضاً مستشـارٌ لكبرى الـشركات في جميع أنحاء العـالم. وتتمثل مجالات اهتمام د. توربان الحالية في نظم دعم القرار على شبكة الإنترنت، والتجارة الاجتماعية، وصُنع القرار بشكل تعاوني.

الفصل الأول

نظرة عامة على ذكاء الأعمال والتحليلات وعلم البيانات

أهداف التعلُّم:

- فهم الحاجة إلى الدعم المحوسب لصنع القرار الإداري.
- التعرُّف على تطوُّر مثل هذا الدعم المحوسب للحالة الراهنة - علم التحليلات / البيانات.
- وصف منهجية ومفاهيم ذكاء الأعمال.
- فهم أنواع مختلفة من التحليلات، والاطلاع على تطبيقات مختارة.
- فهم النظام البيئي للتحليلات لتحديد اللاعبين الأساسيين المتنوعين والفرص الوظيفية.

إنَّ بيئة الأعمال (مناخ الأعمال) تتغير باستمرار حتى أصبحت أكثر تعقيدًا. وتخضع المنظمات، سواءً العامة أو الخاصة، لضغوط تجبرها على سرعة الاستجابة للظروف المتغيرة وأن تكون مبتكرةً في طريقة عملها. وتتطلب هذه الأنشطة من المنظمات أن تكون مرنةً وأن تتخذ قرارات إستراتيجية وتكتيكية وتشغيلية متكررة وسريعة، بعضها معقد جداً. اتخاذ مثل هذه القرارات يتطلب كميات كبـيرة من البيانات والمعلومات والمعارف المرتبطة بالموضوع، والتي يجب أن تتم معالجتها في كثير من الأحيان وفق إطارٍ معين وبشكلٍ عاجل وسريع وعادةً مايتطلب بعض الدعم المحوسب.

ويتناول هذا الكتابُ اسـتخدام تحليلات الأعمال كدعم محوسـب لاتخاذ القرارات الإدارية، ويُركِّز على أسـس ونظريات دعم القرار، وكذلك على الأدوات والتقنيات التجارية المتاحة، ويقدِّم هذا الكتاب أساسـيات تلك التقنيات وطرق بناء الأنظمة واسـتخدامها. ونتبع في ذلك نهج EEE لتقديم هـذه المواضيع: Exposure (التعـرض)، وExperience (التجريـب)، وExploration (الاستكشاف). ويقدِّم هذا الكتاب بشكلٍ أساسي استكشاف تقنيات التحليل المختلفة وتطبيقاتها؛ لتصبح مصدر إلهام للطالب لتعلم كيفية اسـتخدام المنظمات لتلـك التحليلات في اتخاذ قرارات أو ميزات تنافسـية. ونعتقد أن هذا التعرُّض لما يحدث باستخدام التحليلات ونتائجها هو مكون رئيسٌ لتعلُّم تلك التحليلات. لوصف تلك التقنيات سنقدم أمثلةً لأدوات برمجية مُحدَّدة بالإمكان اسـتخدامها لتطويـر هـذه التطبيقات. وفي وصف الأسـاليب، نعرض كذلـك أدوات البرمجيات المحددة التي يمكن استخدامها لتطوير مثل هذه التطبيقات. ولا يقتصر الكتاب على أداة برمجية

واحدة، حتى يتمكَّن الطالب من تجريب هذه التقنيات باستخدام أيِّ أداة من الأدوات البرمجيات المتاحة. ونأمل أن يُمكِّن هذا التعرض والتجريب تحفيز القراء على استكشاف قدرات هذه التقنيات كلٌّ في مجاله. ولتسهيل مثل هذا الاستكشاف؛ فإننا سنورد في هذا الكتاب التدريبات التي توجِّه القارئ إلى شبكة جامعة تيراداتا وغيرها من المواقع، والتي تشمل تدريباتٍ قائمةً على العمل الجماعي إذا لزم الأمر.

يعرضُ هذا الفصل التمهيدي مقدمةً للتحليلات بالإضافة إلى نظرة عامة على هذا الكتاب، ويحتوي على الأجزاء التالية:

١-١ مقال افتتاحي: التحليلات الرياضية - حدود مثيرة لتعلُّم وفهم تطبيقات التحليلات.

١-٢ بيئات الأعمال المتغيرة والاحتياجات المتطورة لدعم القرار والتحليلات.

١-٣ تطوُّر دعم القرار المحوسب لعلم البيانات/ التحليلات.

١-٤ إطار عمل ذكاء الأعمال.

١-٥ نظرة عامة على التحليلات.

١-٦ أمثلة على التحليلات في مجالات مختارة.

١-٧ مقدمة موجزة لتحليلات البيانات الضخمة.

١-٨ نظرة عامة على النظام البيئي للتحليلات.

١-٩ خطة الكتاب.

١-١٠ موارد، وروابط، والاتصال بشبكة جامعة تيراداتا.

١-١ مقال افتتاحي: التحليلات الرياضية - حدود مثيرة لتعلُّم وفهم تطبيقات التحليلات:

إنَّ تطبيق التحليلات على مشكلات الأعمال من بين المهارات الأساسية التي ستتعلمها في هذا الكتاب، ويجري الآن تطبيقُ العديد من هذه الأساليب لتحسين عملية صُنع القرار في جميع جوانب الرياضة، وهو مجالٌ مهم جدًا يُسمَّى التحليلات الرياضية. وتُعَدُّ تحليلات الرياضة بمثابة فن وعلم جمع بيانات عن الرياضيين والفِرق؛ لتكوين رؤى من شأنها تحسينُ القرارات الرياضية، مثل: تحديد اللاعبين الذين سيتمُّ الاستعانة بهم والأجور التي سيتمُّ دفعها لهم، وكيف سيلعبون وكيف سيتمُّ تدريبهم، وكيفية الحفاظ على صحتهم، ومتى يمكن بيعهم ومتى يعتزلون. وبالنسبة للفِرق؛ فإنَّ ذلك ينطوي على قرارات الأعمال، مثل تسعير التذاكر، فضلاً عن القرارات المتعلقة بقائمة اللاعبين وتحليل نقاط قوة وضعف كلِّ منافس، والعديد من القرارات اليومية المتعلقة باللعبة.

في الواقع إنَّ التحليلات في مجال الرياضة لها طابعٌ خاص؛ نظراً لأهمية هذا المجال في كونه مجال أعمال كبيراً يحقق إيرادات تبلغ ١٤٥ مليار دولار أمريكي تقريباً بشكلٍ سنوي؛ بالإضافة إلى ١٠٠ مليار دولار أخرى في الرهانات القانونية و٣٠٠ مليار دولار في الرهانات غير القانونية. في عام ٢٠١٤م وفقاً لمنظمــة Price Waterhouse تــمَّ إنفاق ١٢٥ مليون دولار فقط عــلى التحليلات (أقل من ٠,١٪ من الإيرادات)، ومن المتوقع أن ينمو الإنفاق بمعدل جيد ليصل إلى ٤,٧ مليار دولار بحلول عام ٢٠٢١م[٢].

حظي استخدامُ التحليلات في مجال الرياضة بشعبيةٍ من خلال كتاب Moneyball الذي كتبه Lewis Michael في عام ٢٠٠٣م، وكذلك فيلم بطولة Brad Pitt في عام ٢٠١١م والذي عرَض فيه Billy Beane المدير العام لأوكلاند أ استخدام البيانات والتحليلات؛ لتحويل فريق خاسر إلى فريق فائز. على وجه الخصوص؛ قام بيلي بن بتعيين محلل اســتخدم التحليلات لصياغة لاعبين قادرين على الحصول على قاعدة بدلاً من اللاعبين الذين تميزوا في التدابير التقليدية مثل عمليات الضرب أو القواعد المسروقة. هذه الأفكار مكَّنتهم من صياغة توقعات تغفل عنها الفِرق الأخرى برواتب مبدئية معقولة؛ لقد نجحت تلك الأفكار ووصلوا إلى التصفيات في عامي ٢٠٠٢ و٢٠٠٣م.

والآن يتــمُّ اســتخدامُ التحليلات في جميع مناحي الرياضة. ويمكِن تقســيمُ التحليلات ما بين المكتــب الأمامي والمكتب الخلفي، ويُظهِر مقال اســتبيان Tom Davenport وصفاً جيداً لثلاثين مثالاً[٣]. وتشمل تحليلاتُ أعمال المكتب الأمامي تحليل سلوك الجماهير بدءًا من النماذج التنبؤية لتجديد تذاكر الموســم ومبيعات التذاكر العادية، إلى توقعات التهديف من قِبل المشــجعين فيما يتعلَّــق بالفريق والرياضيين والمدربين، والمُلّاك. وهذا يشــبه إلى حَدٍّ كبــيرٍ إدارة علاقات العملاء التقليديــة (CRM). ويُعَــدُّ التحليلُ المالي كذلك من المجالات الرئيســة؛ حيث يُعد الحدُّ الأقصى للراتب أو حدود البعثة جزءًا من المعادلة.

ويشملُ استخدام المكاتب الخلفية تحليلَ كلٍّ من الرياضيين الأفراد، وكذلك لعب الفريق. وبالنسبة للاعبين الفرديين؛ يتمثل التركيزُ على نماذج التوظيف وتحليلات الكشافة وتحليلات القوة واللياقة البدنية

(1) Source: Changing the Game: Outlook for the Global Sports Market to 2015, Price Waterhouse Coopers Report, appears at https://www.pwc.com/gx/en/hospitality-leisure/pdf/changing-the game-outlook-for-the-global-sports-marketto-2015.pdf. Betting data from https://www.capcredit.com/how-much-americansspend-on-sports-each-year/.

(2) Sports Analytics Market Worth $4.7B by 2021," Wintergreen Research Press Release, covered by PR Newswire at http://www.prnewswire.com/news-releases/sports-analytics-market-worth-47-billion-by-2021-509869871.html, June 25, 2015.

وكذلـك التنمية، ومتابعــة الأداء PMs؛ لتجنُّب الإفراط في التدريب والإصابات، وكذلك فإن ارتجاج المخ مــن المجالات المهمة في الرياضة. وتشــمل تحليلات الفريق الإســتراتيجيات والتكتيــكات، والتقييمات التنافســية، والخيــارات المُثلى للقائمة في إطارٍ مختلــف من الحالات الميدانيــة أو في الملاعب. وتوضِّح الأمثلةُ التوضيحية التالية كيف تستخدم ثلاثُ منظمات رياضية البيانات والتحليلات لتحسين العمليات الرياضية، بنفس الطريقة التي قامت بها التحليلات بتحسين صناعة القرارات التقليدية في الصناعة.

مثال ١- مكتب الأعمال:

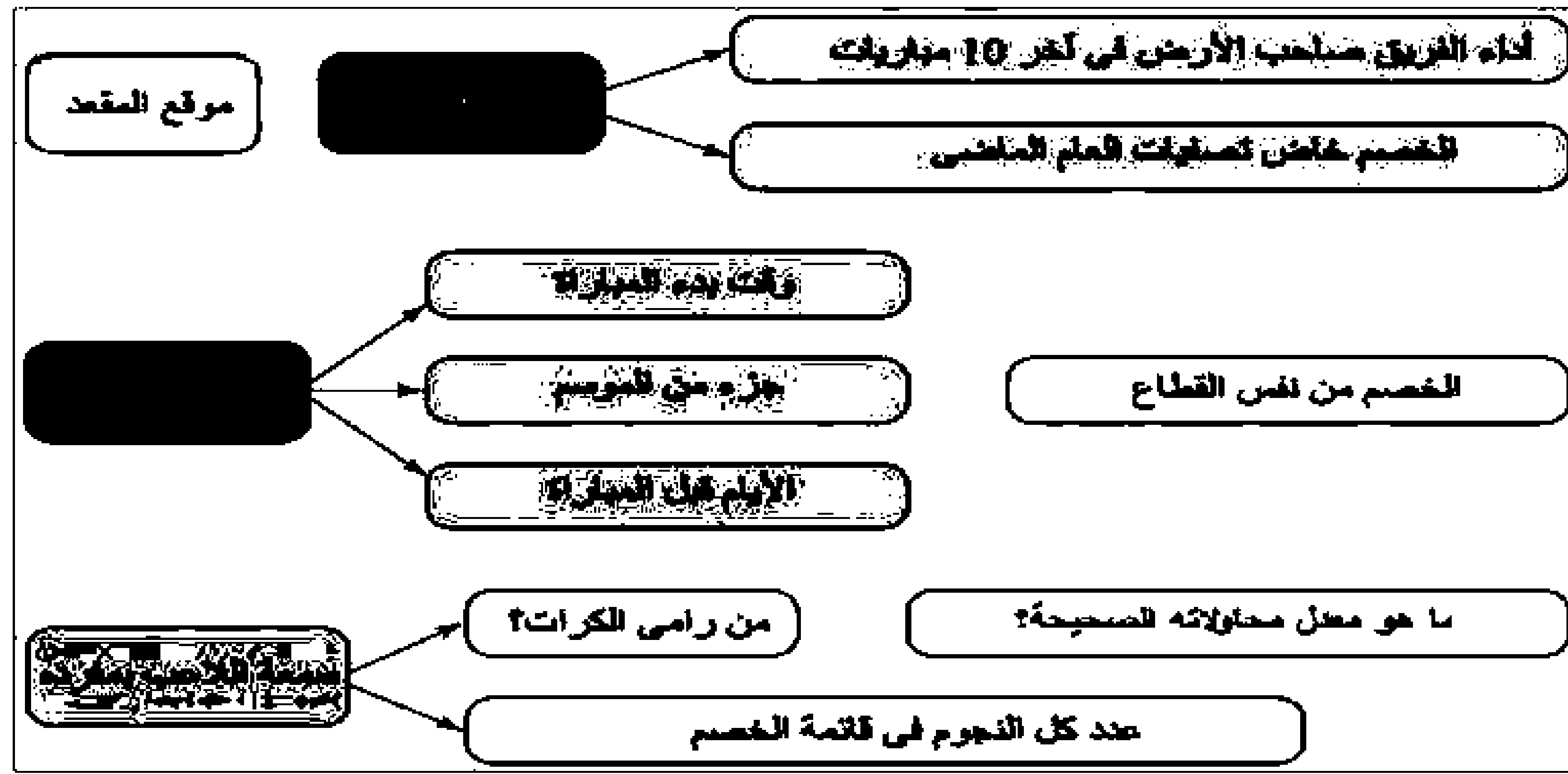
يعمــلُ ديف وارد محللَ أعمال لفريقٍ كبير بلعبة البيســبول للمحترفين، ويركِّز عمله على العوائد وتحليــل مبيعات التذاكر ســواءً تذاكر موســم كامل أو تذكــرة واحدة. بعض الأســئلة التي تخضع لمســؤوليته تتضمَّن لماذا يُجدِّد (أو لا يُجدِّد) حاملو التذاكر الموســمية تذاكرهم، وكذلك العواملُ التي تدفع الجمهور لشراء تذكرة مقعد فردي في آخر لحظة، وأيضاً كيفية تســعير التذاكر. وتشــمل بعض الأساليب التحليلية التي يستخدمها مكتب Dave إحصاءات بسيطة عن سلوك الجماهير، مثل الحضور العام والإجابات عن أسئلة الاستبيان حول احتمال الشراء مرةً أخرى. ومع ذلك؛ فقد يختلف ما يقوله الجماهــير عمَّا يفعلونه. ويقوم Dave بإجراء اســتبيان المشــجعين من خــلال المكان المحجوز لمقعد التذكرة «الدرجة» ويسأل عن احتمال تجديد تذاكر الموسم. ولكنه عندما يقارن ما يقولونه مقابل ما يفعلونه؛ يكتشف اختلافات كبيرة (انظر الشكل ١-١). حيث تبيَّن له أن ٦٩٪ من المشجعين في مقاعد الدرجة الأولى الذين قالوا في الاســتبيان أنهم «قد لا يُجدِّدون» قد جدَّدوا بالفعل، وكوَّن هذا بالنسبة له رؤيةً جيدةً مفادها أنَّ العملاء في الخلايا الخضراء ســيقومون على الأرجح بتجديد تذاكرهم؛ لذلك يحتاجون إلى جهد أقل في التسويق وصرف الدولارات مقارنةً بالعملاء في الخلايا الزرقاء.

الطبقة	محتمل جدا	محتمل	ربما	محتمل لا	مؤكد لا
1	92	88	75	60	45
2	88	81	70	55	38
3	80	76	68	55	36
4	77	72	65	45	25
5	75	70	60	35	25

شكل ١-١: تجديد تذاكر الموسم - نتائج الدراسة الاستقصائية

(3) Sources: Thomas Davenport, "Analytics in Sports: The New Science of Winning," International Institute for Analytics White paper, sponsored by SAS, February 2014. On the SAS Web site at: http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/iia-analytics-in-sports-106993.pdf. (Accessed July 2016).

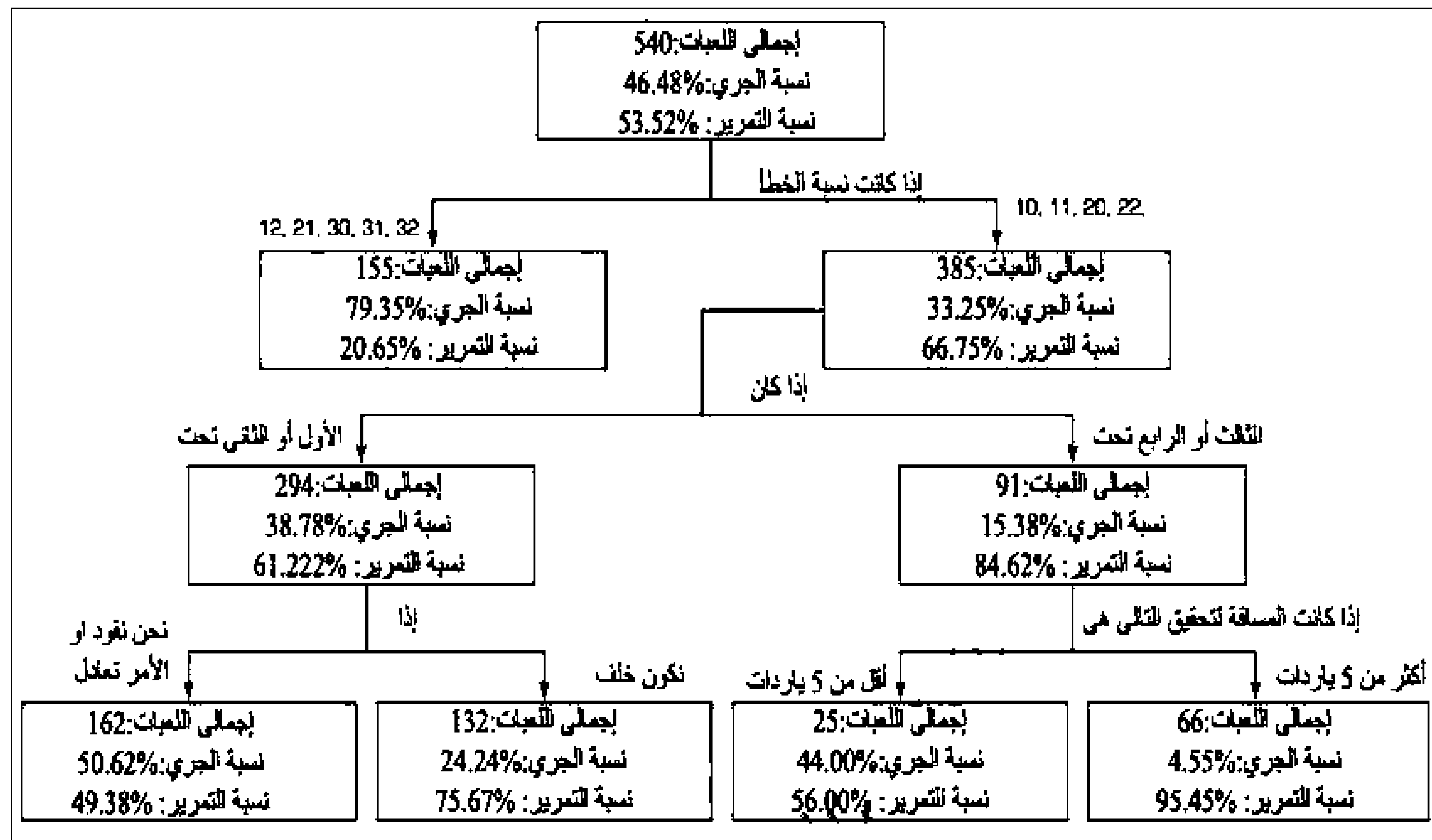
ولكن هناك عِدة عوامل تؤثرُ على سـلوك الجمهور فيما يتعلق بشراء التذاكر وخاصةً السعر؛ مما يستدعي إجراء إحصاءات أكثر تطوُّراً وتحليل بيانات. في كلا المجالين وخصوصاً تذاكر اللعبة الواحدة؛ يقود ديف اسـتخدام التسـعير الديناميكي؛ لينقل المجال من التسـعير الثابت البسيط المعتمد على موقع المقعد إلى التسـعير الديناميكي اليومي المتأثـر بالصعود والهبوط. يُعتبَر هذا المجال مجالًا بحثيًّا ثريًّا للعديد من الفِرق الرياضية من شأنه زيادة فرص تحسين إيراداتهم. على سـبيل المثال: عند التسعير يجبُ مراعاة سِـجل الفريق ومن يلعبون وتواريخ وتوقيت المباريات التي يلعب فيها اللاعبون النجوم لكلِّ فريق، وتاريخ تجديد تذاكر الموسـم أو شراء تذاكر فردية. بالإضافة إلى عوامل أخرى مثل مكان المقعد، وعدد المقاعد، والمعلومات التي تواكب وقت المباراة، مثل الازدحام المروري وقت المباراة وحتى الطقس. انظر الشكل ١-٢.



شكل ١-٢: التسعير الحركي للعمل السابق - دوري البيسبول الأساسي

أيٌّ من هذه العوامل أكثر أهميةً؟ وما مقدار أهميته؟ اعتماداً على الخبرة الإحصائية الواسعة؛ قام ديف ببناء نماذج انحدار لاسـتخلاص العوامل الرئيسـية التي تقودُ تلك السلوكيات التاريخية وقام بإنشـاء PMs لتحديد كيفية إنفاق موارد التسـويق لدفع الإيرادات. كما قام بإنشاء نماذج مُخصَّصة لحملة التذاكر الموسمية لتحديد شرائح العملاء الذين سيقومون أو لن يقوموا بالتجديد، أو مَنْ لا يسـتطيعون اتخاذ قرار مناسـب ومن خلال ذلك يمكن القيامُ بحملات تسـويقية أكثر دقـةً. وبالإضافة إلى ذلك؛ يقوم بإجراء إحصـاء للتوقعات المتعلقة بالتهديف من خلال تعليقات الجمهور، مثل التغريدات التي تساعده على تقسيم المشجعين إلى شرائح ولائية مختلفة. وتساعد الدراسات الأخرى عن دوافع حضور لعبة فردية إدارة التسويق على فهم تأثير الهدايا مثل القمصان الرياضية، أو اقتراحات حول أماكن نشر الإعلانات التلفزيونية الفورية.

وبالإضافة إلى الإيرادات؛ يوجد العديدُ من المجالات التحليلية الأخرى التي يعمل عليها فريق Dave، بمــا في ذلك التســويق، وعوائد إعلانــات الراديو والتليفزيون وتقديــم البيانات إلى المدير العام بشأن المفاوضات على الراتب، ومشاريع التحليلات الخاصة بالنظر إلى الحَدِّ الأقصى للراتب، وفعالية الترويج بما في ذلك القنوات الإعلانية، والتوعية بالعلامة التجارية، وكذلك تحليلات الشركاء.



شكل ١-٣: شجرة القرارات المتتالية لتنفيذ أو تمرير اللعبات

مثال٢- المدرب The Coach:

بوب بريدلوف مدربُ كرة قدم لفريق إحدى الكليات الكبيرة، ويُركِّز بشــكلٍ كاملٍ على الفوز بالمباريــات. يركِّز بوب على اســتقطاب أفضل اللاعبــين في المدارس الثانويــة، وتطوير مهاراتهم لتتناســب مع أنظمــة هجومه ودفاعه، وكذلــك الحصولُ على أقصى جهدٍ ممكــن منهم في أثناء المباريات. يقع تحت مســؤولياته أيضاً الإجابة عن أسئلة مثل: مَنْ سنستقطبُ من اللاعبين؟ ما التدريبات التي تســاعدُ على تطوير مهاراتهم؟ ما مدى صعوبة دفع وتحفيز لاعبينا؟ وما نقاط قوة وضعف المنافسين؟ وكيف يمكننا معرفة أو التعرُّف على اتجاه لعبهم؟

ولحُسْــن الحظ؛ اســتعان فريقه بخبيرةِ عمليــات الفريق الجديدة دار برنــاك، المتخصِّصة في مساعدة المدربين في اتخاذ القرارات التكتيكية، وتعمل دار مع فريقٍ من الطلاب المتدربين الذين يقومون بإجراء تحليلات للخصم باستخدام فيلم المباراة، والذي يقوم المدربُ بالتعليق عليه لإنشاء

نموذج شجرة القرارات المتتالية كما في الشكل ١-٣ للتنبؤ بما إذا كانت المباراة التالية ستكون مباراة جري أو تمرير. أمَّا بالنسبة للمنسق الدفاعي؛ فقد قاموا ببناء خرائط حرارية كما في الشكل ١-٤ لكلِّ هجوم للخصم لتوضيح مَيْلهم إلى التمرير إلى اليسار أو اليمين وإلى مناطق التغطية الدفاعية. وأخيراً؛ قاموا بإجراء تحليلات لبعضِ السلاسل الزمانية كما في الشكل ١-٥ على المباريات القوية (تُعرف على أنها الحصول على ١٦ تمريرة صحيحة أو الجري لمسافة ١٢ ياردة أو أكثر). لكلِّ مباراة، يقارنون نتائج التحليل بتشكيلاتهم الدفاعية وهجوم الفِرق المنافسة لمساعدة المدرب بريدلوف على التفاعل بسرعة أكبر مع التحوُّلات في أثناء المباراة، سيتم شرح هذه الأساليب التحليلية بشكلٍ مفصل في الفصول الثاني والخامس والسابع.

الهجوم

A	B	C
تام:35	تام:6	تام:22
اجمالي:46	اجمالي:8	اجمالي:27
76.8%	75.00%	81.48%
ساخن :4	ساخن :5	ساخن :2

الخط العشوائي

1	2	3	4	5
تام:25	تام:12	تام:14	تام:8	تام:25
اجمالي:35	اجمالي:24	اجمالي:28	اجمالي:14	اجمالي:44
71.4%	50%	50%	57.14%	56.81%
ساخن :1	ساخن:0	ساخن :0	ساخن :0	ساخن :1

6	7	8	9
تام:7	تام:13	تام:7	تام:15
اجمالي:10	اجمالي:21	اجمالي:10	اجمالي:27
70%	61.9%	70%	55.55%
ساخن :2	ساخن :9	ساخن :6	ساخن:8

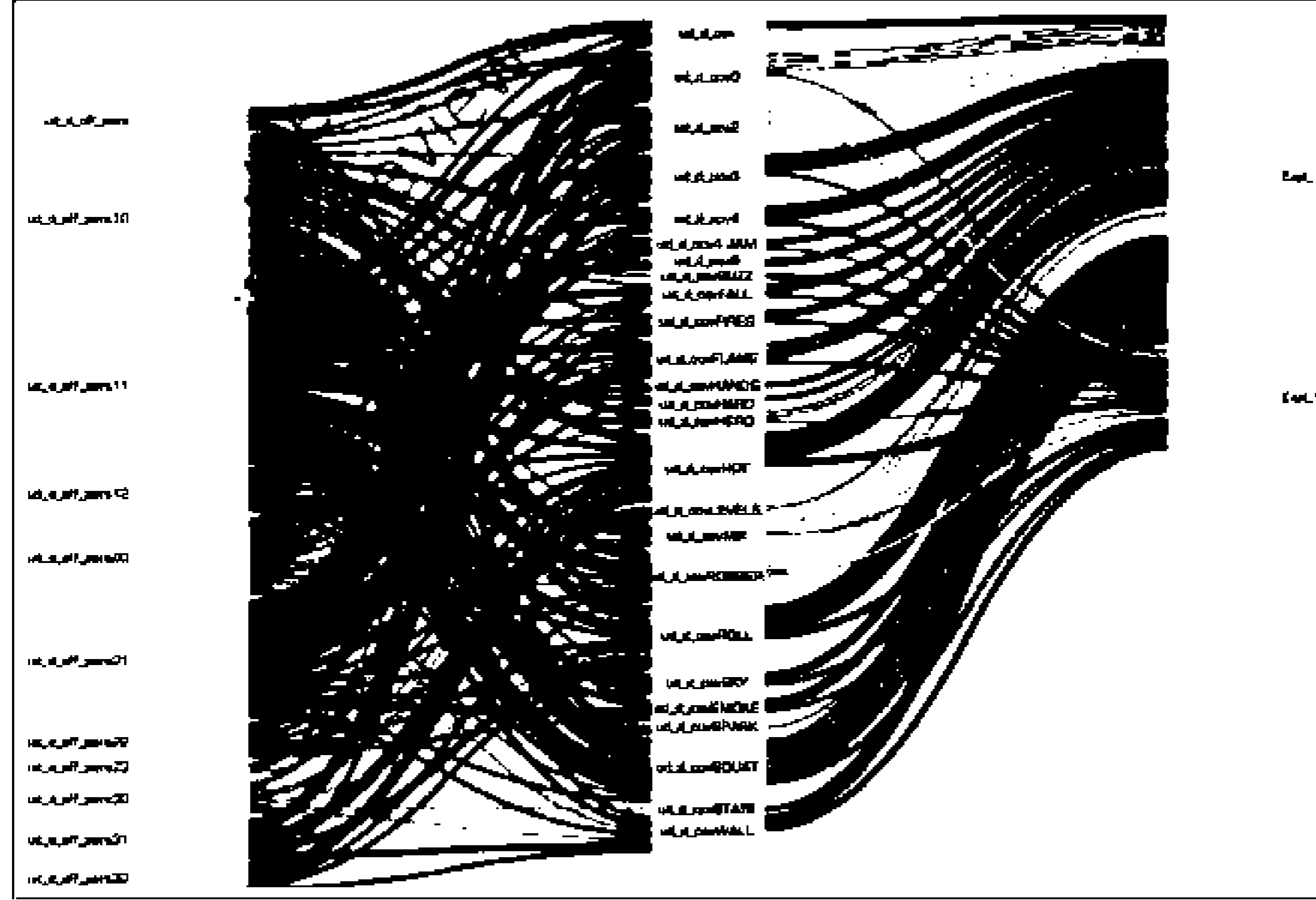
X	Y	Z
تام:1	تام:7	تام:5
اجمالي:13	اجمالي:18	اجمالي:15
7.69%	38.88%	33.33%
ساخن :1	ساخن :7	ساخن :6

الدفاع

شكل ١-٤: تحليل منطقة الخريطة الحرارية للتمريرات

وتشمَل الأعمال الجديدة التي تقوم بها Dar بناء نماذج أفضل لاستقطاب الرياضيين من المدارس الثانوية. فعلى سبيل المثال؛ يقدِّم فريق العمل في كلِّ عامٍ مِنحاً دراسيةً لثلاثة طلاب يتمُّ اختيارهم للفريق. أما بالنسبة لـ Dar؛ فإن اختيار أفضل اللاعبين لا يقتصر على المقاييس البسيطة مثل مدى سرعة اللاعبين، ومدى ارتفاع قفزاتهم، وطول أذرعهم؛ بل يتخطاها إلى معايير جديدة، مثل مدى سرعة تدوير رؤوسهم لالتقاط التمريرات، وما أنواع وأوقات رد الفعل للاعب

عـلى المؤثرات المتعددة، ومدى دقة تنفيذه للتمريـرات. وتُنشر بعض أفكار دار في توضيح هذه المفاهيم على موقع TUN على شبكة الإنترنت؛ ابحث عن BSI Case of Precision Football[٤].



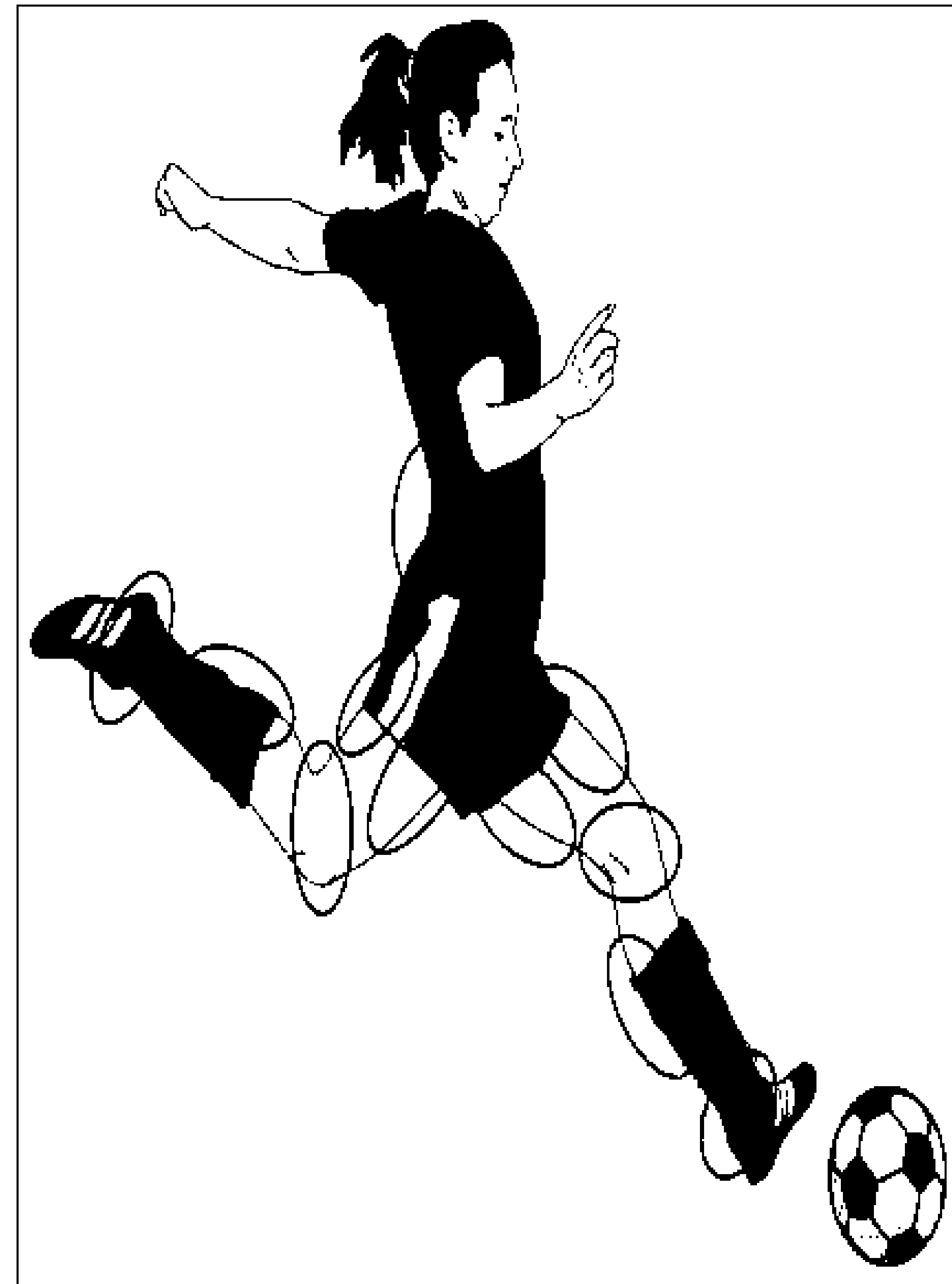
شكل ١-٥: تحليل السلاسل الزمنية للعبات الساخنة

مثال٣- المدرب The Trainer:

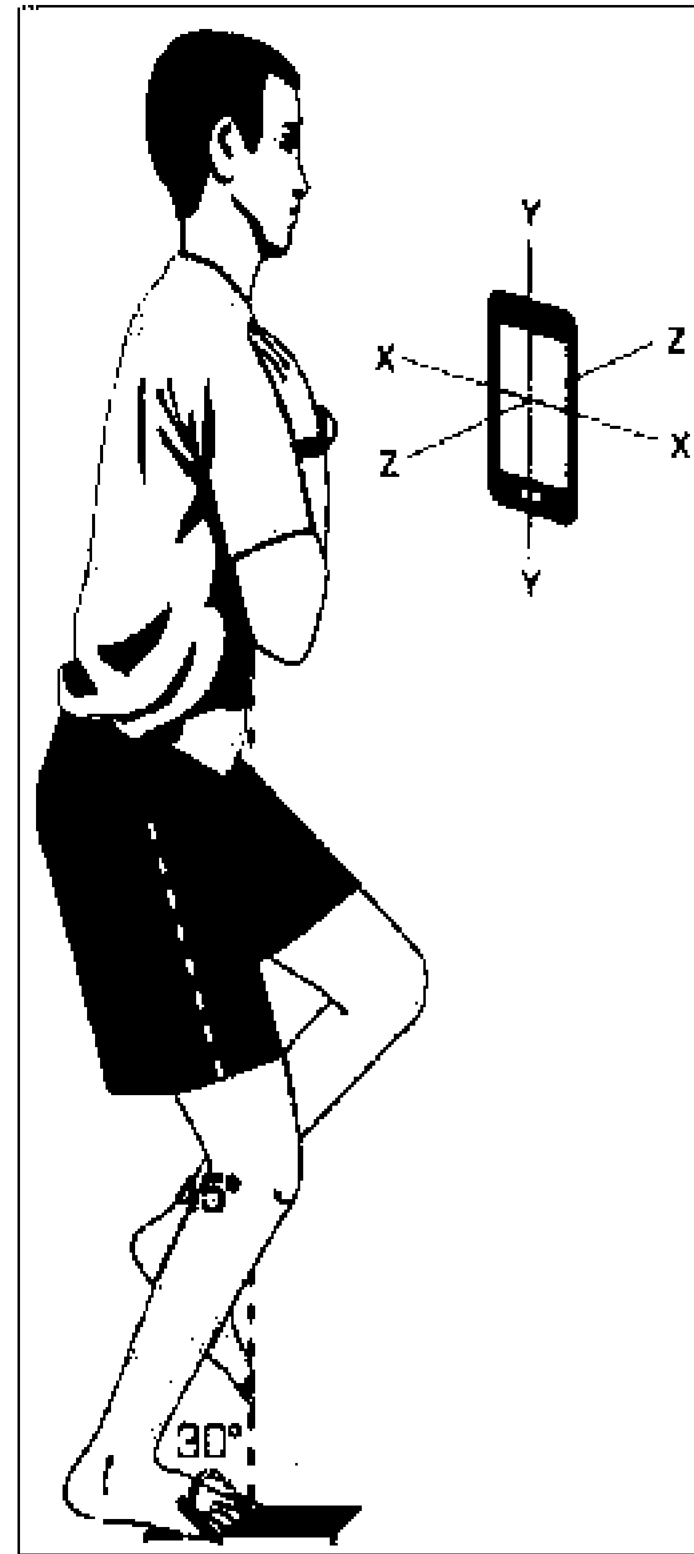
يعمل الدكتور دان جونســون مدربـاً لفريق كرة القدم بكلية المرأة لمسـاعدة اللاعبات على الحِفـاظ عـلى صحتهـن وتقديم المشـورة للمدربين بخصوص قـدرة التحمُّل لـكلِّ لاعبة خلال التدريبات. كذلك يهتمُّ جونســون برفاهية اللاعبات بمـا في ذلك مقدار النوم ووقت الراحة لكلِّ لاعبة بين فترات التدريب الثقيلة والخفيفة، والهدف من ذلك التأكُّد من استعداد اللاعبات للعب بأقصى قدرٍ ممكنٍ من الكفاءة في أثناء المباريات.

ولحُسْـن الحظِّ، وبسبب اســتخدام الأجهزة الرياضية المتطوِّرة؛ كان هناك المزيدُ من البيانات التي مكَّنَت الدكتور Dan من التحليل. فقد كانت اللاعبات يستخدمن سُتْرات تحتوي على أجهزة الاستشـعار يمكن من خلالها قياسُ الأحمال الداخلية؛ مثل ضربات القلب، ودرجة حرارة الجسم، ومُعدلات التنفس علاوةً على سُتْرات التسارع التي تقيسُ الأحمال الخارجية، مثل العدو لمسافات والسرعـات وكذلك التسـارع والتباطؤ. ومن ثم التعرُّف على اللاعبات الـلاتي يُعطِين أقصى جهدٍ خلال التدريبات ومن لَسْنَ كذلك.

وينصبُّ تركيزُ جونســـون حاليًّا على الأبحاث التي تساعده في التنبؤ بإصابات اللاعبات وكيفية الوقايــة منها (الشــكل ١-٦). ويمكن أن توفِّر تدريبات بســيطة مثل الوقوف على ســاق واحدة وجلوس القرفصاء واختبار إمساك ساق ثم السَّاق الأخرى مع وجود فرق نقاط أكثر من ١٠٪ رؤى مفيدة لجونســون حول نقاط قوة وضعف الجســم كما في شــكل ١-٧. فإذا أُصِيبَت لاعبة خلال المباراة، يمكن للمدرب إجراءُ اختبار جانبي لرد فعل تحفيزي على جهاز محمول، والتي تضيفُ إلى بروتوكولات الصدمات التقليدية. وتظهر أجهزة استشــعار النوم مَنْ الذي يحتاج إلى الحصول على راحة كافية (أو الذين سهروا طوال الليل). كما استعان د. دان بمختبر التصوير بالرنين المغناطيسي في الحرم الجامعي للقيام بمسح الدماغ الدوري لإظهار أيٍّ الرياضيين عُرضةً للإصابة بإصابات الدماغ.



شكل ١-٦: نماذج إصابات كرة القدم[٥]



شكل ١-٧: اختبار قرفصاء الساق الواحدة - اختبار قوة الجسم الأساسية

(٥) "إصابات كرة القدم للسيدات،" المركز القومي لتقارير بحوث الإصابات الرياضية الكارثية، NCAA. يتم إنتاج أوراق واقع الإصابات الرياضية NCAA بواسطة مركز Datalys لأبحاث الإصابات الرياضية والوقاية منها بالتعاون مع الرابطة القومية لرياضة الجامعات وإصابات رياضة STOP. تصدر في:

https://www.ncaa.org/sites/default/files/NCAA_W_Soccer_Injuries_WEB.pdf

(تم الوصول إليها في نوفمبر ٢٠١٦)

أسئلة على هذه الأمثلة:

١- ما العوامل الثلاثة التي قد تكون جزءًا من PM لتجديد تذكرة الموسم؟

٢- ما الأساليب التي يمكن أن تستخدمها فِرق كرة القدم للقيام بتحليل الخصم؟

٣- كيف يمكِن للأجهزة القابلة للارتداء تحسين صحة وسلامة اللاعب؟ وما أنواع التحليلات الجديدة التي يمكن أن يستخدمها المدربون؟

٤- ما استخدامات التحليلات الأخرى التي يمكنك تصورها في الألعاب الرياضية؟

ماذا يمكننا أن نتعلَّم من هذه المقالة القصيرة:

بعيدًا عن مُحلِّلي الأعمال في المكاتب الأمامية والمدربين وخبراء الأداء؛ فإن هناك العديدَ من الأشخاص الآخرين في الألعاب الرياضية يستخدمون البيانات، بدءًا من لاعبي الجولف الذين يقيسون ظروف التربة والأرض لبطولات PGA وحكام البيسبول وكرة السلة الذين يتم تقييمهم بناءً على قراراتهم الصحيحة وغير الصحيحة التي يجرونها. وفي الواقع؛ يصعب العثور على مجال من مجالات الرياضة لا يتأثر بتوفُّر المزيد من البيانات، وخاصةً من أجهزة الاستشعار.

إنَّ المهارات التي ستتعلمها في هذا الكتاب لتحليل الأعمال تنطبق على الرياضة. فإذا كنتَ ترغب في التعمق في هذا المجال؛ فإننا نشجعك على الاطلاع على قسم تحليلات الرياضة بشبكة جامعة تيراداتا (TUN)، وهي مصدرٌ مجاني للطلاب وأعضاء هيئة التدريس. وفي هذا الموقع؛ سوف تجد وصفاً لما يجب قراءته لمعرفة المزيد عن التحليلات الرياضية، ومجموعات من الأماكن التي يمكنك العثور بها على مجموعات البيانات المتاحة للجمهور للتحليل؛ فضلًا عن أمثلة لمشاريع الطلاب في تحليلات الرياضة ومقابلات المهنيين الرياضيين الذين يستخدمون البيانات والتحليلات للقيام بوظائفهم.

١-٢ بيئات الأعمال المتغيرة والاحتياجات المتطوِّرة لدعم القرار والتحليلات:

توضِّح المقالة الافتتاحية كيف يمكن للصناعة بأكملها أن تستخدم تحليلات لوضع تقارير عمَّا يحدث، والتنبؤ بما يمكن أن يحدث، ومن ثم اتخاذ القرارات أيضاً لتحقيق أفضل استفادة من

(4) Sources: Business Scenario Investigation BSI: The Case of Precision Football (video). (Fall 2015). Appears on http://www.teradatauniversitynetwork.com/About-Us/Whats-New/BSI-Sports-Analytics-Precision-Football//, Fall 2015. (Accessed September 2016).

الوضع القائم. وتتطلَّب هذه الخطوات من المنظمة جمع وتحليل المخزون الضخم من البيانات. ومن الاستخدامات التقليدية في وظائف كشوف المرتبات وإمساك الدفاتر، توغلت النظم المحوسبة الآن في مجالات إدارية معقدة تتراوح بين تصميم وإدارة المصانع الآلية إلى تطبيق أساليب تحليلية لتقييم عمليات الاندماج والاستحواذ المقترحة. ويعرف جميع المديرين التنفيذيين تقريبًا أن تقنية المعلومات ذات أهمية حيوية لأعمالهم ويستخدمون تقنيات المعلومات على نطاق واسع.

وقد انتقلت تطبيقات الحاسوب من أنشطة معالجة المعاملات ورصدها إلى تحليل المشكلات وتطبيقات الحلول، وتمَّ القيامُ بالكثير من الأنشطة باستخدام التقنيات المستندة إلى السحابة، التي تم الوصول إليها في كثيرٍ من الحالات من خلال الأجهزة المتنقلة. إن أدوات التحليل وذكاء الأعمال مثل مستودع البيانات، والتنقيب في البيانات، والمعالجة التحليلية الفورية (OLAP) ولوحات المعلومات، واستخدام الأنظمة القائمة على السحابة لدعم القرار؛ هي أهمُّ عناصر الإدارة الحديثة في الوقت الحاضر. وأصبح من المحتمل أن يكون لدى المديرين أنظمة معلومات وشبكات (سلكية أو لاسلكية) عالية السرعة لمساعدتهم في مهمتهم الأكثر أهميةً وهي اتخاذ القرارات. وفي كثيرٍ من الحالات، يتمُّ تنفيذ مثل هذه القرارات آلياً بشكل روتيني؛ مما يلغي الحاجة إلى أيِّ تدخل إداري.

وبالإضافة إلى الزيادة الواضحة في قدرات المكوِّنات المادية والبرمجيات والشبكات؛ ساهمت بعض التطورات بشكلٍ واضح في تسهيل زيادة دعم القرارات والتحاليل بعدة طرق، منها:

- **التواصل والتعاون الجماعي:** يتمُّ اتخاذ العديد من القرارات اليوم من قِبل المجموعات التي قد يكون أعضاؤها في مواقع مختلفة. وتستطيع المجموعات التعاون والتواصل بسهولة باستخدام أدوات التعاون وكذلك الهواتف الذكية في كلِّ مكان. ويكتسب التعاون أهميةً خاصة على امتداد سلاسل التوريد؛ إذ يجبُ على الشركاء تبادل المعلومات مع الموِّردين والعمال بصفة مستمرة. وقد يكون تجميع مجموعة من صانعي القرار، وخاصةً الخبراء، في مكانٍ واحد باهظ التكلفة. وبالتالي يمكن لنظم المعلومات أن تحسِّن عملية التعاون في مجموعةٍ ما وتمكِّن أعضاءها من التواجد في مواقع مختلفة (لتوفير تكاليف السفر). والأهم من ذلك؛ أن هذا التعاون في سلاسل التوريد يسمح للمصنِّعين بمعرفة أنماط التغير في الطلب في الوقت المناسب، وبالتالي يتفاعلون مع تغيُّرات السوق بشكل أسرع.

(٥) «إصابات كرة القدم للسيدات» المركز القومي لتقارير بحوث الإصابات الرياضية الكارثية، NCAA. يتم إنتاج أوراق واقع الإصابات الرياضية NCAA بواسطة مركز Datalys لأبحاث الإصابات الرياضية والوقاية منها بالتعاون مع الرابطة القومية لرياضة الجامعات وإصابات رياضة STOP. تصدر في: https://www.ncaa.org/sites/default/files/NCAA_W_Soccer_Injuries_WEB.pdf. (تم الوصول إليها في نوفمبر ٢٠١٦).

- **تحســـين إدارة البيانات:** تنطوي العديدُ من القرارات على حســـابات مُعقدة تعتمد على بيانات مخزنة في مستودعات مختلفة في أماكن داخل وربما خارج المنظمة. وقد تتضمَّن البيانات النص والصوت والرســـومات والفيديو، وقد تكون هـــذه الصيغ بلغات مختلفة. وفي كثيرٍ من الأحيان يكـــون مـــن الضروري نقل البيانات بسرعـــة من مواقع بعيدة. ومن خـــلال أنظمة اليوم يمكن البحث عن البيانات اللازمة وتخزينها ونقلها بسرعة وبتكلفة اقتصادية وأمان وشفافية.
- **إدارة مستودعات البيانات العملاقة والبيانات الضخمة:** تحتوي مستودعات البيانات الضخمة، مثل تلك التي تديرها Walmart، على كميات هائلة من البيانات. وتتوفَّر أساليب خاصة، بما في ذلك الحوســـبة المتوازية، وHadoop/Spark ومثيلاتها لتنظيم البيانات والبحث فيها والتخلص منهـــا. وتتناقص تكاليف تخزين والتنقيب في البيانات بسرعة، هذه التقنيات تندرجُ تحت فئةٍ واســـعة من البيانات الضخمة التي تمَّ جمعها من مصادر متنوعة وبأشكال مختلفة، جمع تلك البيانات يعطي رؤية كاملة للأداء التنظيمي هذه الرؤية لم تكن ممكنةً في الماضي.
- **الدعـــم التحليـــلي:** ومع المزيدِ من تقنيات البيانات والتحليل، يمكـــن تقييمُ المزيد من البدائل، ويمكن تحســـين التنبـــؤات، ويمكن إجراء تحليـــل المخاطر بسرعة، كما يمكن جمـــع آراء الخبراء (وبعضهـــم في مواقع نائية) بسرعة وبتكلفة منخفضة. ويمكن اســـتخلاص الخبرة حتى بشـــكلٍ مباشر من النظم التحليلية. وباســـتخدام هذه الأدوات، يســـتطيعُ صانعو القرار إجراء محاكاة معقـــدة، والتحقق من العديـــدِ من الســـيناريوهات المحتملة، وتقييم الآثـــار المتنوعة بسرعة واقتصادية. وهذا بالطبع هو محور عدة فصول في هذا الكتاب.
- **التغلُّـــب على الحدود المعرفية في معالجة وتخزيـــن المعلومات:** وفقاً لـ Simon (1977)؛ إذ إنَّ للعقل البشري قـــدرةً محدودةً على معالجة وتخزين المعلومات. ويجد الناسُ أحياناً صعوبةً في استرجاع واســـتخدام المعلومات دون أخطاء بسبب حدودهم المعرفية. ويشير مصطلحُ الحدود المعرفيـــة إلى أن قدرة الفرد على حلِّ المشـــكلات تكون محدودةً عندمـــا تكون هناك حاجة إلى مجموعة واســـعة من المعلومات والمعرفة المتنوعة. فالنظم المحوســـبة تمكِّن الناس من التغلُّب على حدودهم الإدراكية من خلال الوصولِ السريع ومعالجة كميات هائلة من المعلومات المخزنة.
- **إدارة المعرفة:** وقد جمعت المنظماتُ مستودعاتٍ هائلةً من المعلومات عن عملياتها الخاصة، والعملاء، والإجراءات الداخلية، وتفاعلات الموظفين، وما شابه ذلك. من خلال الاتصالات المهيكلة وغير المهيكلة التـــي تجري بين مختلف أصحـــاب المصالح. وأصبحت نظم إدارة المعرفة بالنســـبة للمديرين مصادر للدعم الرسمي وغير الرسمي لاتخاذ القرارات، وإن كانت لا تسمى أحياناً بنظم إدارة المعرفة. التقنيات مثل تحليل النصوص وIBM Watson جعلت من الممكن استخراج قيمة من هذه المستودعات.

- **الدعم في أيِّ مكانٍ وفي أيِّ وقت:** وباستخدام التقنية اللاسلكية، يمكن للمديرين الوصول إلى المعلومات في أيِّ وقتٍ ومن أيِّ مكان، وتحليل وتفسير ذلك، والتواصل مع المعنيين. وربما كان هذا أكبر تغيير حدث في السنوات القليلة الماضية. فالسرعة التي تحتاجُ إلى معالجة المعلومات وتحويلها إلى قرارات قد غيرت بالفعل توقعات المستهلكين والشركات على حدٍّ سواء. وقد أدَّى كلُّ ذلك إلى استخدام الدعم المحوسب للقرارات منذ أواخر الستينيات، وبصفة خاصة منذ منتصف التسعينيات. وقد أتاح نمو التقنيات المتنقلة ومنصات وسائل التواصل الاجتماعي والأدوات التحليلية مستوًى مختلفاً من دعم نظم المعلومات (IS) للمديرين. ولا يمتد هذا النمو في توفير الدعم المعتمد على البيانات لأي قرارٍ إلى المديرين فحسب ولكن أيضاً إلى المستهلكين. وسنقوم أولاً بدراسة لمحة عامة عن التقنيات التي تمَّت الإشارة إليها على نطاق واسع باسم ذكاء الأعمال. من هذا المنطلق سوف نقوم بتوسيع آفاقنا لتقديم أنواع مختلفة من التحليلات.

أسئلة مراجعة على القسم ١-٢:

١- اذكر بعض الاتجاهات الرئيسة الموجهة بالنظام والتي عزَّزَت عملية صُنع القرار التي تدعمها نظم المعلومات "IS" إلى مستوى جديد؟

٢- اذكر بعض قدرات نظم المعلومات التي يمكن أن تسهِّل اتخاذ القرارات الإدارية.

٣- كيف يمكن للحاسوب المساعدة في التغلُّب على الحدود الإدراكية للإنسان؟

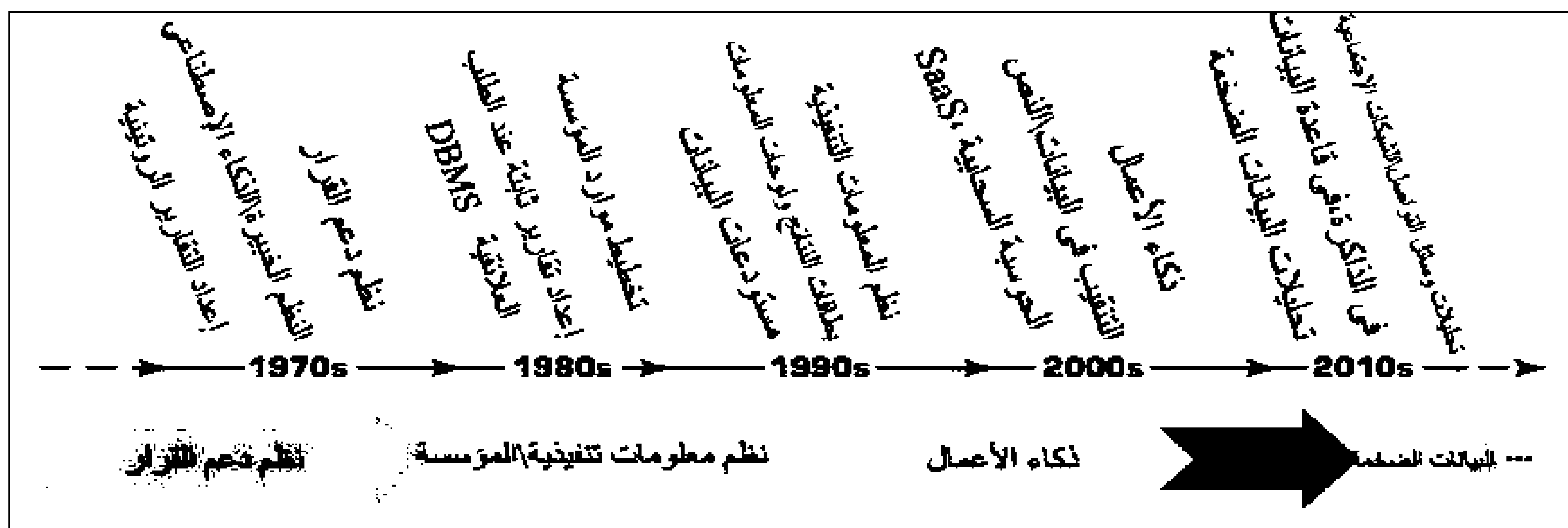
١-٣ تطوُّر دعم القرار المحوسب لعلم البيانات/ التحليلات:

يُبيِّن الجدول الزَّمني الوارد في الشكل (١-٨) المصطلحات المستخدمة لوصف التحليلات منذ السبعينيات. وخلال السبعينيات؛ كان التركيزُ الأساسي لنظم المعلومات في دعم القرارات على تقديم تقارير دورية منظمة يمكِن للمدير استخدامُها في اتخاذ القرارات (أو تجاهلها). وقد بدأت الأعمال في إنشاء تقارير روتينية لإبلاغ صانعي القرار (المديرين) بما حدث خلال فترة سابقة (يوم،

المصدر والاعتمادات: لقد قام الدكتور Dave Schrader بتلك المساهمة، والذي تقاعد بعد ٢٤ عاماً من التطوير والتسويق المتقدمين في تيراداتا، وقد ظل في مجلس المستشارين بشبكة جامعة تيراداتا؛ ليقضي فترة تقاعده في مساعدة الطلاب وأعضاء هيئة التدريس على معرفة المزيد عن التحليلات الرياضية. تم تصميم صور كرة القدم (الأشكال ١-٣ و١-٥) بواسطة طلاب الدراسات العليا Peter Liang وJacob Pearson بجامعة أوكلاهوما الحكومية، كجزء من مشروع طلابي في ربيع عام ٢٠١٦م. وتمَّ اقتباسُ صور التدريب (الأشكال ١-٦ و١-٧) من الصور التي قدَّمها كلٌّ من البروفيسور Gary Wilkerson من جامعة تينيسي في تشاتانوجا والبروفيسور Ashish Gupta من جامعة أوبورن.

أو أسـبوع، أو شهر، أو ربع سنة). وعلى الرغم من أنه كان من المفيدِ معرفة ما حدث في الماضي؛ فـإنَّ المديرين كانوا بحاجةٍ إلى أكثر من ذلك؛ لأنهم يحتاجـون إلى مجموعةٍ متنوعة من التقارير عـلى مسـتويات مختلفة من التفصيل لفهـم ومعالجة الاحتياجات والتحَدِّيـات المتغيرة للشركة بشـكلٍ أفضل. وعادةً ما تُسـمَّى هذه بنُظُم المعلومات الإدارية (MIS). في أوائل السبعينيات، وقـد طرح Scott-Morton المفاهيم الرئيسـة لنظام دعم القـرار DSS. وقد عُرفت نظم دعم القرار DSSs بأنها «أنظمة حاسوبية تفاعلية، تساعد صُناع القرار على استخدام البيانات والنماذج لحلِّ المشـكلات غير المهيكلة» Gorry و(1971) Scott-Morton. وفيما يلي تعريف كلاسـيكي آخر لـ DSS، والذي قدَّمه Keen و(1978) Scott-Morton:

> تجمع نظم دعم القرار بين الموارد الفكرية للأفراد مع قدرات الحاسب؛ لتحسين نوعية القرارات. وهو نظام دعمٍ قائمٍ على الحاسوب لصناع القرار الإداريين الذين يتعاملون مع المشكلات شبه الهيكلية.



شكل ١-٨: تطور دعم القرار وذكاء الأعمال والتحليلات

لاحظ أنَّ مصطلح «نظام دعم القرار»، مثل نظام المعلومات الإدارية والعديد من المصطلحات الأخـرى في مجال تقنية المعلومات، هو تعبيرٌ خالٍ من المحتوى (بمعنى أنه يعني أشـياء مختلفة لأشخاص مختلفين). ولذلك؛ لا يُوجَد تعريف مقبول عالميًا لنظام دعم القرار.

وخـلال الأيـام الأولى من التحليـلات؛ غالبًا ما يتـمُّ الحصول على البيانات مـن خبراء المجال باسـتخدام العمليات اليدوية (أيْ: المقابلات والاسـتقصاءات) لبناء النماذج الرياضية أو المعرفة القائمة على حَلِّ مشكلات التحسين المقيَّدة. وكانت الفكرة هي تحقيق الأفضل بالموارد المحدودة. وعادةً ما تُسـمَّى نمـاذج دعم القرار هذه باسـم بحوث العمليـات (OR). ولقد تمَّت معالجة المشـكلات التي كانت معقدةً إلى درجة لم يكن يمكن حلها على النحو الأمثل (باستخدام أساليب

البرمجة الرياضية الخطية أو غير الخطية) باستخدام أساليب الاستدلال مثل نماذج المحاكاة. (سنقدِّم هذه التحاليل الوصفية لاحقاً في هذا الفصل وبتفاصيل أكثر في الفصل السادس).

وفي أواخر السبعينيات وأوائل الثمانينيات؛ وبالإضافة إلى نماذج بحوث العمليات الناضجة التي كانت تُستَخدَم في العديد من الصناعات والنظم الحكومية؛ ظهرت مجموعةٌ جديدةٌ ومثيرةٌ من النماذج، مثل: نظم الخبراء القائمة على القواعد. وقد بشَّرت هذه الأنظمة بحصر معارف الخبراء في شكل تستطيع الحواسيب معالجته (من خلال مجموعة من القواعد أو الاستدلالات)؛ بحيث يمكنُ استخدام هذه القواعد للتشاور بنفس الطريقة التي يستخدمها خبراءُ المجال لتحديد مشكلة مهيكلة ووصف الحلِّ الأكثر احتمالًا. وقد سمحت النظم الخبيرة ESs بتوفير الخبرة النادرة عندما يلزم الأمر؛ وذلك باستخدام نظام دعم القرار DSS «الذكي».

ولقد شَهِدَت الثمانينيات تغيرًا كبيرًا في طريقة استحواذ المنظمات على البيانات المتعلقة بالأعمال التجارية. وكانت الممارسة القديمة تتمثل في وجود عدة نظم معلومات مُفكَّكة مُصمَّمة خصيصًا لتجميع بيانات المعاملات لمختلف الوحدات أو الوظائف التنظيمية (مثل: المحاسبة، والتسويق، والمبيعات، والتمويل، والتصنيع). وفي الثمانينيات؛ تمَّ دمجُ هذه الأنظمة كنظم معلومات على مستوى المؤسسة التي نسميها الآن نظم تخطيط موارد المؤسسة (ERP). وتمَّ استبدالُ مخططات تمثيل البيانات القديمة المتسلسلة وغير المعيارية بأنظمة إدارة قواعد البيانات العلائقية (RDBM). وقد أتاحت هذه النظم تحسين الحصول على البيانات وتخزينها، فضلًا عن العلاقات بين حقول البيانات التنظيمية مع انخفاض ملحوظ في تكرار المعلومات.

وبرزت الحاجةُ إلى نظم إدارة قواعد البيانات العلائقية (RDBM) ونظم تخطيط موارد المؤسسة (ERP) عندما أصبحت سلامة البيانات واتساقها مشكلةً؛ مما أعاق بشكلٍ كبيرٍ فعالية الممارسات التجارية. وبتخطيط موارد المؤسسات؛ يتمُّ جمعُ كلِّ البيانات من جميع أركان المؤسسة ودمجها في مخطط ثابتٍ؛ بحيث يكون لكلِّ جزءٍ من المنظمة حقُّ الوصول إلى نسخة واحدة من الحقيقة متى لزم الأمر. وبالإضافة إلى ظهور أنظمة تخطيط موارد المؤسسات، أو ربما بسبب هذه الأنظمة؛ أصبح إعداد تقارير الأعمال ممارسة تجارية حسب الطلب، أو حسب الحاجة. ويمكن لصناع القرار أن يُقرِّروا متى يحتاجون أو يريدون إنشاء تقارير متخصصة لتقصِّي المشكلات والفرص التنظيمية.

وفي التسعينيات، أدَّت الحاجة إلى إعداد تقارير أكثر تنوعًا إلى تطوير أنظمة المعلومات التنفيذية (تمَّ تطويرُ نظم المعلومات التنفيذية EISs ونظام دعم القرار خصيصًا للمديرين التنفيذيين واحتياجاتهم من اتخاذ القرار) وقد صُمِّمت هذه الأنظمة على شكلٍ لوحات بيانية

وبطاقــات قياس؛ بحيــث يمكن أن تكون بمثابة عروض جذّابة بصريًا مع التركيز على أهمِّ العناصر التــي يمكــن لصانعي القرار من خلالها تتبُّع مؤشرات الأداء الرئيســة. ولجعــل هذا التقرير أكثر تنوعًا مع الحفاظ على سلامة المعاملات لنظم المعلومات التجارية؛ كان من الضروري إنشاء طبقةِ البيانات المتوســطة المعروفة باسم مســتودعات البيانات (DW) كمستودع لدعم تقارير الأعمال وصُنــع القرار تحديــدًا. وخلال وقتٍ قصير جــدًا؛ اعتمدت معظمُ الشركات الكبيرة والمتوســطة الحجم مستودعَ البيانات كمنصة لصُنع القرار على مستوى المؤسسة. وحصلت لوحاتُ المعلومات وبطاقــات الأداء على بياناتها من مســتودعات البيانات (DW)؛ وبذلك تحسَّــنَت كفاءة أنظمة المعاملات التجارية التي يُشار إليها في معظمها بأنظمة تخطيط موارد المؤسسة (ERP).

وفي العِقدِ الأول من القرن العشرين؛ بدأ يُطلَق على الأنظمة القائمة على مستودعات البيانات اسم «أنظمة ذكاء الأعمال». ومع ازدياد كمية البيانات الطولية المتراكمة في مستودعات البيانات؛ كان لا بُدَّ أن تواكب قــدرات الأجهزة والبرمجيات الاحتياجات المتطورة والسريعة التغيُّر لصانعي القرار. وبســبب السوق التنافسية المعولمة؛ يحتاج صانعو القرار إلى استيعاب المعلومات الحالية في قالب ســهل الفهم لمعالجة مشــكلات الأعمال والاستفادة من فرص السوق في الوقت المناسب. وفي حــين يتمُّ تحديث البيانات في مســتودعات البيانات بشــكلٍ دوري؛ فإنهــا لا تعكس أحدث المعلومــات. وللتغلُّــب على هذه المشــكلة المتعلقة بوقت الوصول إلى المعلومــات؛ طوَّر مُورِّدو مســتودعات البيانات نظامًا لتحديث البيانات بشكلٍ أكثر تواترًا؛ مما أدَّى إلى الاستيداع في الوقت المناسب للبيانات، والذي يختلف عن المنهج السابق في التخزين من خلال اعتماد سياسة تحديث البيانــات على أســاس التحديث المطلوب لعناصر البيانات (أيْ: ليــس كلُّ عناصر البيانات تحتاج إلى تحديث بشــكلٍ فوري). إنَّ مُســتودعَات البيانات ضخمةٌ جداً وهي ميزة هائلة. وأصبح من الضروري «استخراج» البيانات الخاصة بالشركة؛ من أجل «اكتشاف» المعلومات الجديدة والمفيدة لتحسين عمليات وممارسات الأعمال، ومن ثم مُصطلحَيْ التنقيب في البيانات والتنقيب في النص.

ومــع تزايدِ حجم البيانــات وأصنافها؛ ظهرت الحاجة إلى المزيدِ مــن التخزين وقوة المعالجة. وعلى الرغم من أنَّ الشركات الكبيرة لديها الوســائل اللازمة لمعالجة هذه المشــكلة؛ فإنَّ الشركات الصغــيرة الحجم تحتاج إلى نمــاذج أعمال أكثر قابليةً للإدارة المالية. وقــد أدَّت هذه الحاجة إلى نماذج الخدمات الموجَّهة نحو الخدمة والبرامج والبنية التحتية كخدمة تحليل الخدمات. وبالتالي؛ تمكَّنَت الشركات الصغيرة من الوصول إلى إمكانيات التحليلات حسب الحاجة، وتدفع فقط مقابل ما استخدموه، بدلاً من الاستثمار المالي في موارد الأجهزة والبرامج الحاسوبية.

ومنذ عام ٢٠١٠م وحتى الآن، ونحن نشهد نقلةً نوعيةً أخرى في طريقة رَصْد البيانات واستخدامها. وبسبب انتشارِ استخدام الإنترنت على نطاقٍ واسع؛ ظهرت وسائلُ جديدة لتوليد البيانات مثل بطاقات

تعريف تــرددات الراديو [RFID]، وعدادات الطاقة الرقمية، وســجلات النقر عبر الإنترنت، والأجهزة المنزلية الذكية، ومُعدَّات المراقبة الصحية القابلة للارتداء، وربما كان أكثرها أهميةً الشبكات الاجتماعية والتــي أدَّت إلى ظهور بيانات غــير مهيكلة، هذه البيانات غنية بالمحتــوى؛ ولكنها صعبة التحليل من ناحية البرمجيات والأجهزة على حَدٍّ ســواء. وفي الآونة الأخيرة؛ ظهر مصطلح البيانات الضخمة ليسلِّط الضوء على التحَدِّيات التي جلبتها لنا هذه البيانات الجديدة. وقد تمَّ تطويرُ العديد من التحســينات في كلٍّ من الأجهزة (فعلى سبيل المثال: المعالجة المتوازية على نِطاق واسع مع الذاكرة الحسابية الكبيرة جدًّا ونظم الحوســبة متعددة المعالجات المتوازية للغاية) والبرمجيات/ الخوارزميات، مثل (Hadoop مع MapReduce وNoSQL) والتي تمَّ تطويرها للتغلُّب على تحديات البيانات الضخمة.

ومــن الصعب التنبؤ بما ســوف يحدث خلال العقــد القادم في مجال التحليــلات، وما المصطلحات الجديدة التي ســتظهر؛ لقد قلَّص الزمــنُ الفجوة بين نماذج أنظمة المعلومات والتحليلات، وسيســتمر في المســتقبل القريب رغم أنَّ التحليلات ليســت جديدةً؛ ولكن شعبيتها تزايدت بشكلٍ ملحوظ مؤخرًا وبفضــل التزايــد الهائل في البيانــات الضخمة؛ أصبحت طــرق جمع وتخزين هــذه البيانات، وأدوات البرمجيــات البديهية، والرؤى التي تعتمد على البيانات أكثر ســهولةً لرجال الأعمال من أيِّ وقتٍ مضى. لذلك، وفي خضم المنافسة العالمية، أصبحت الفرصة كبيرةً لاتخاذ قرارات إدارية أفضل باستخدام البيانات والتحليــلات لزيادة الإيرادات مع خفــض التكاليف من خلال إخراج منتجات أفضل، وتحســين تجربة العملاء، واكتشــاف الاحتيال قبل حدوثه، وتحسين تفاعل العملاء من خلال الاستهداف والإنتاج حسب طلــب العميل مع قوة التحليــلات والبيانات. ويتزايد عددُ الشركات التي تقوم حالياً بتدريب موظفيها على أدوات تحليل الأعمال؛ من أجل تحقيق أكثر فعالية وكفاءة في عمليات صنع القرارات اليومية.

ويركِّــز القســم التالي على إطار عمل ذكاء الأعمال. وعلى الرغم مــن أن معظم الناس يتفقون على أن ذكاء الأعمال قد تطور إلى تحليلات وعلم البيانات؛ فإن العديد من الموردِّين والباحثين لا يزالون يستخدمون هذا المصطلح. لذلك يولي القســم ١-٤ اهتمامًا بهذا التاريخ من خلال التركيز على ما يُســمَّى ذكاء الأعمال. وبعد القسم التالي، نقدِّم التحليلات وسوف نستخدم ذلك كعلامة لتصنيف جميع المفاهيم ذات الصلة.

أسئلة مراجعة على القسم ١-٣:

١- اذكر ثلاثة من المصطلحات التي سبقت التحليلات.

٢- ما الفرق الأســاسي بين النُّظُم التي تُســمَّى نظم المعلومات الإداريــة ونظام دعم القرار ونظم المعلومات التنفيذية؟

٣- هل تطوَّر نظامُ دعم القرار إلى ذكاء أعمال أو العكس؟

١-٤ إطار عمل ذكاء الأعمال:

لقد تمَّ تنفيذُ مفاهيم دعم القرار الواردة في القسمين ١-٢ و١-٣ بشكلٍ تدريجي، تحت أسماء مختلفــة من قِبل العديد من المورِّدين الذين قاموا بإنشــاء أدوات وأســاليب لدعم القرار. وكما لُوحِظ في القسم ١-٣، ومع نموِّ النظم على نطاق المؤسسة؛ تمكَّن المديرون من الوصول إلى تقارير ســهلة الاســتخدام تمكِّنهم من اتخاذ القرارات بسرعة. وبدأت هذه الأنظمة، التي كانت تُسمَّى عمومًــا نظم المعلومــات التنفيذية EISs، تقدم تصورًا إضافيًــا، وتنبيهات، وقدرات لقياس الأداء. وبحلول عام ٢٠٠٦م؛ ظهرت المنتجات والخدمات التجارية الرئيسة تحت مصطلح ذكاء الأعمال.

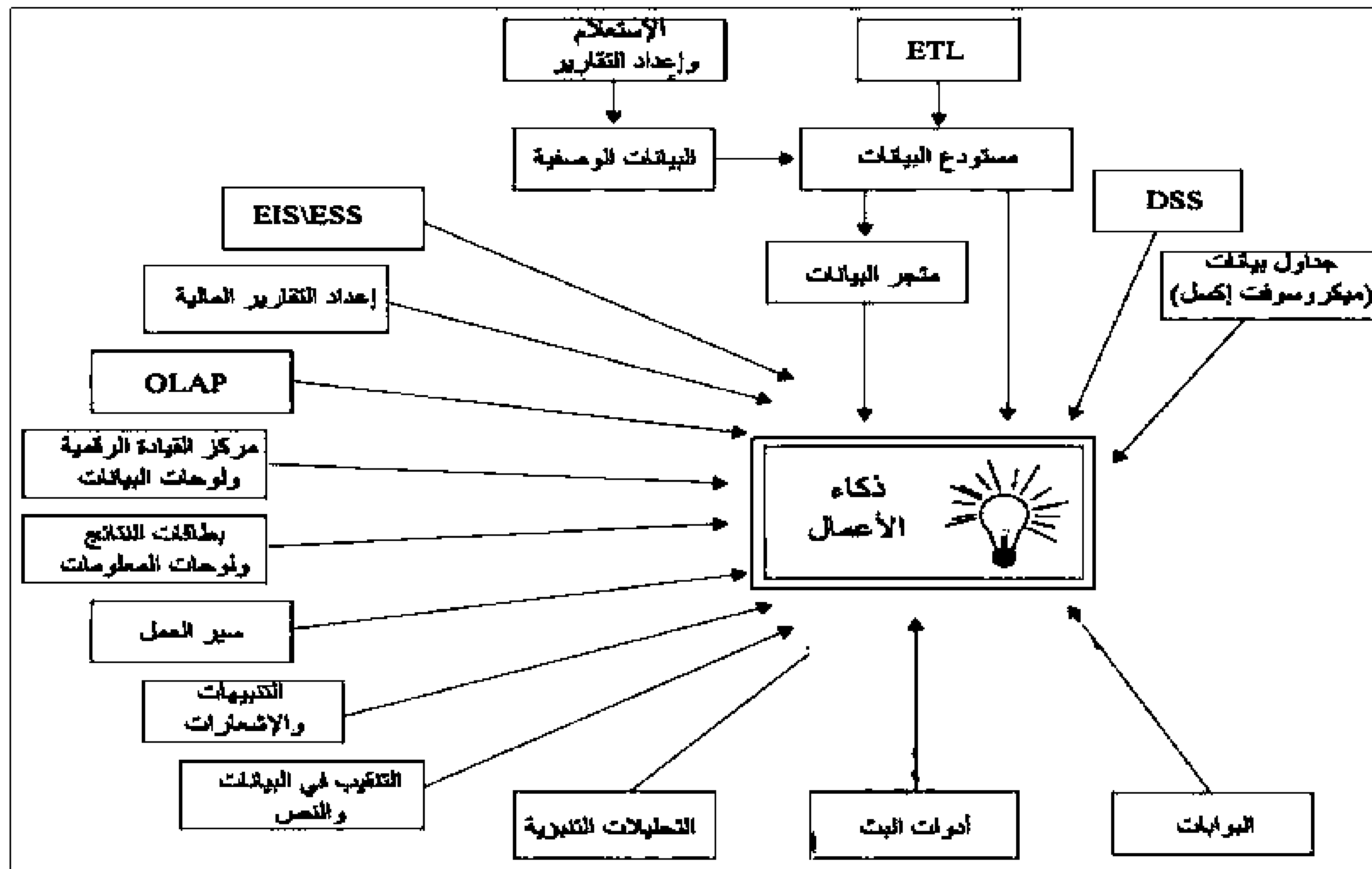
تعريفات ذكاء الأعمال:

ذكاء الأعــمال (BI) هو مصطلحٌ شــامل يجمع بــين البنيات والأدوات وقواعــد البيانات والأدوات التحليليــة والتطبيقات والمنهجيــات. ويتزايد عددُ الشركات التي تقوم حاليــاً بتدريب موظفيها على أدوات تحليــل الأعمال؛ من أجل تحقيق أكثر فعالية وكفاءة في عمليات صُنع القرارات اليومية وهناك نــوعٌ من الارتباك حول ذكاء الأعمال ناتج مــن موجة الاختصارات والكلمات الطنانة المرتبطة به (على سبيل المثال: إدارة أداء الأعمال [BPM]). فالهدف الرئيس لذكاء الأعمال؛ هو تمكين الوصول التفاعلي وفي بعض الأحيان بشكلٍ فوري إلى البيانات، لتمكين معالجة البيانات، وإعطاء مديري الأعمال والمحلِّلين القــدرة على إجراء التحليلات المناســبة. ومن خلال تحليــل البيانات التاريخيــة والحالية، والمواقف، والعروض، يستطيع صُناع القرار الحصول على رؤى قيِّمة تمكِّنهم من اتخاذ قرارات أفضل وأكثر استنارة. وتستند عملية ذكاء الأعمال إلى تحويل البيانات إلى معلومات، ثم إلى قرارات، وأخيرًا إلى إجراءات.

لمحة تاريخية عن ذكاء الأعمال:

لقد صِيغَ مصطلحُ ذكاء الأعمال من قِبل مجموعة Gartner في منتصف التســعينيات ومع ذلك، وكما يشيرُ التاريخ في القسم السابق؛ فإنَّ هذا المفهوم أقدم من ذلك بكثير، وله جذوره في نظم تقارير MIS في السبعينيات. وخلال تلك الفترة؛ كانت نظم التقارير ثابتة، وكانت ثنائية الأبعاد، ولم تكن لديها قدرات تحليلية. وفي أوائل الثمانينيات؛ ظهر مفهوم نظم المعلومات التنفيذية EISs، وقد وســع هذا المفهوم الدعم المحوســب لدى كبار المديرين والمديرين التنفيذيين. ومن بين الإمكانات التي قدَّمتها التقارير الديناميكية المتعددة الأبعاد (المخصصة أو حسب الطلب) التنبؤ وتحليل الاتجاهات، وتحليل التفاصيــل، والوصول إلى الحالة، وعوامل النجــاح الحرجة. وقد ظهرت هذه الميزات في العشرات من

المنتجات التجارية حتى منتصف التســعينيات، ثم ظهرت نفس القدرات وبعض الإمكانات الجديدة تحت اســم ذكاء الأعــمال. واليوم؛ يحتــوي ذكاء الأعمال الجيد القائم على أســاس نظام المعلومات المؤسسية على جميع المعلومات التي يحتاجها المسؤولون التنفيذيون في مجال المعلومات. لذلك، تم تحويل المفهــوم الأصلي لنظام المعلومات التنفيذي EIS إلى ذكاء الأعمال. وبحلول عام ٢٠٠٥؛ بدأت أنظمة ذكاء الأعمال لتشمل قدرات الذكاء الاصطناعي بجانب قدرات تحليلية قوية. ويُوضِّح الشكل ١-٩ مختلف الأدوات والأســاليب التي يمكن إدراجها في نظام ذكاء الأعمال، ويوضح كذلك تطور ذكاء الأعمال، وتبين الأدوات المعروضة في الشــكل قــدرات ذكاء الأعمال وتتوفر هذه القدرات في منتجات ذكاء الأعمال الأكثر تطورًا. وبعضها الآخر متخصص في بعض منها فقط.



شكل ١-٩: تطور ذكاء الأعمال (Bi)

بنية ذكاء الأعمال:

تتكوَّن بنية نظام ذكاء الأعمال من أربعةِ مكوِّنات رئيسة هي:

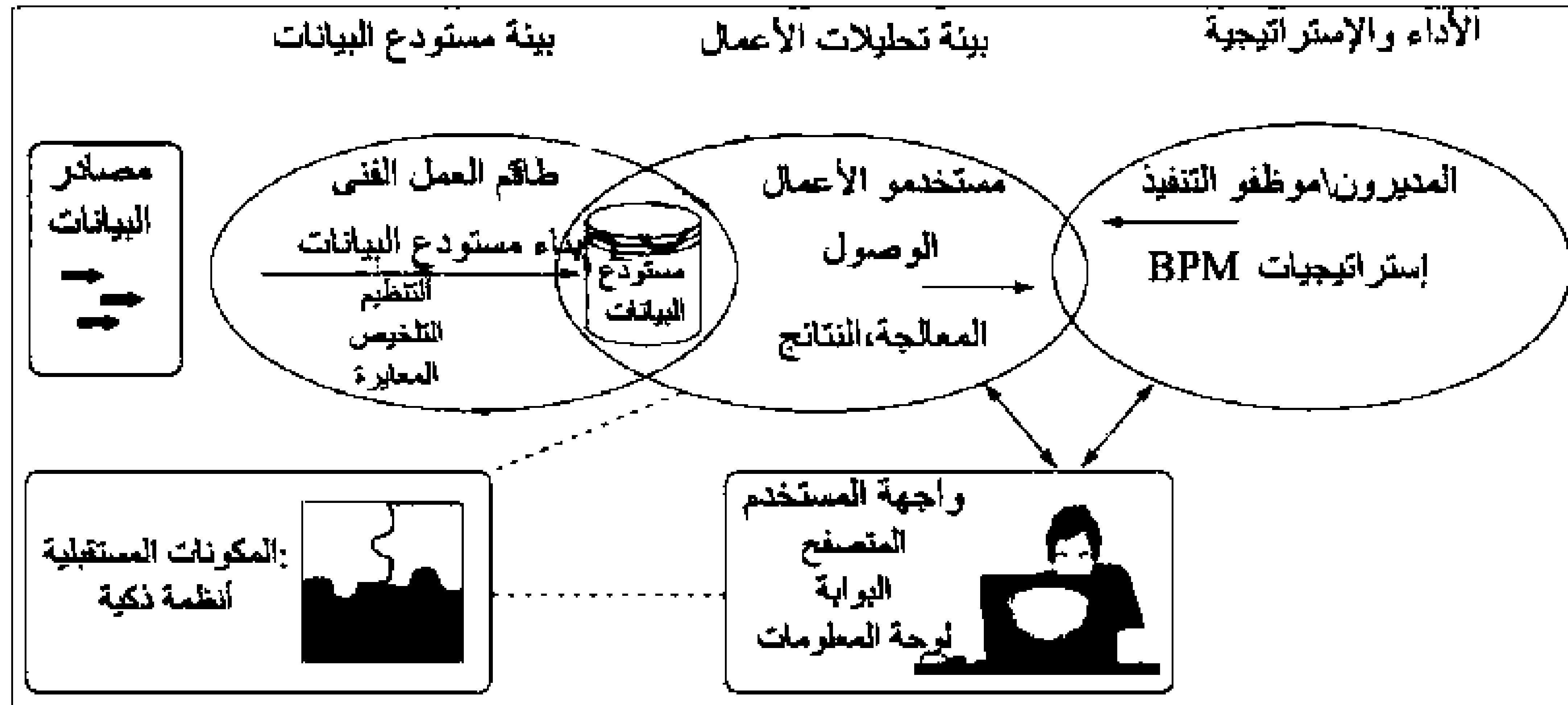
١- مستودعات البيانات مع بيانات مصدرها.

٢- تحليــلات الأعــمال، وهــي مجموعــة أدوات معالجة وتنقيــب وتحليل البيانــات الموجودة في مستودعات البيانات.

٣- نظام إدارة الأعمال لمراقبة وتحليل الأداء.

٤- واجهة مستخدم (على سبيل المثال: لوحة المعلومات).

ويوضح الشكل ١-١٠ العلاقة بين هذه المكونات.



source: based on w.eckerson smart companies in the 21st century the secrets of creating. successful business intelligent sollutions. the data warehousing inistilute.seattle . wa.2003..p.32.iiiuslration 5

شكل ١-١٠: بنية عالية المستوى لذكاء الأعمال

أصول ومحركات ذكاء الأعمال:

من أين أتت الطرق الحديثة لمستودع البيانات وذكاء الأعمال؟ وما هي جذورها؟ وكيف تؤثر تلك الجذور على الطريقة التي تدير بها المنظمات هذه المبادرات اليوم؟ إن الاستثمارات الحالية في تقنية المعلومات تخضع لمزيدٍ من التدقيق من حيث تأثيرها على خط الأساس واحتماليتها. وينطبق الشيء نفسه على مستودعات البيانات وتطبيقات ذكاء الأعمال التي تجعل هذه المبادرات ممكنة.

وتضطر المنظمات إلى استيعاب بياناتها وفهمها وتسخيرها لدعم القرارات لتحسين العمليات التجارية. فالقوانين والتنظيمات (مثل: قانون ساربينز - أوكسلي لعام ٢٠٠٢) تتطلب الآن من سير الأعمال توثيق عملياتهم التجارية والتوقيع على قانونية المعلومات التي يعتمدون عليها وتقديم تقرير إلى حاملي الأسهم. وعلاوةً على ذلك؛ أصبحت أوقات دورة الأعمال مضغوطةً للغاية وأسرع وأكثر استنارةً وأصبحت هناك حاجة لاتخاذ قرار أفضل، وبالتالي، زاد احتياجُ المديرين إلى المعلومات الصحيحة في الوقت المناسب وفي المكان المناسب. وهذا هو شعار النهج الحديث لذكاء الأعمال.

ويجب على المنظمات أن تعمل بذكاء مع الاهتمام البالغ بإدارة مبادرات ذكاء الأعمال التي تُعَدُّ جانبًا ضروريًا من جوانب الأعمال. وليس مثيرًا للدهشة أن تناصر المنظمات بشكلٍ متزايد ذكاء الأعمال في تجسيدها الجديد كتحليلات. وتعرض حالة عملية ١-١ أحد تطبيقات

ذكاء الأعمال الذي ساعد العديد من شركات الطيران، وكذلك بالطبع، الشركات التي تقدم هذه الخدمات لشركات الطيران.

حالة عملية ١-١

شركة Sabre تساعد عملاءها من خلال لوحات المعلومات والتحليلات

تُعَدُّ Sabre واحدةً من الشركات العالمية الرائدة في صناعة الرحلات السياحية. وتقوم الشركة بتوفير الخدمات للأشخاص والشركات. وهي تخدمُ المسافرين ووكلاء السفر والشركات ومورِّدي خدمة السفر من خلال أربع شركات رئيسة هي: Travelocity، وSabre Travel Network، وSabre Airline Solutions، وSabre Hospitality Solutions. وتشكِّل البيئة الاقتصادية العالمية المتقلبة الحالية تحديات تنافسية كبيرة لصناعة الطيران. وللمُضيّ قُدمًا في المنافسة، اعترفت شركة Sabre Airline Solutions بأن المديرين التنفيذيين في شركات الطيران يحتاجون إلى أدوات مُحسّنة لإدارة قراراتهم التجارية والتخلُّص من الإجراءات التقليدية واليدوية التي تستغرق وقتًا طويلًا لجمع المعلومات المالية وغيرها من المعلومات اللازمة للمبادرات القابلة للتطبيق. وهذا يتيح دعم القرار في الوقت المناسب لشركات الطيران في جميع أنحاء العالم لتحقيق أقصى قدرٍ من العائدات على المعلومات من خلال أفكار قيادية وذكاء قابل للتطبيق، وقيمة للعملاء من البيانات المتنامية.

قامت شركة Sabre بتطوير مستودع بيانات سفريات المؤسسة (ETDW) باستخدام تيراداتا للاحتفاظ ببيانات الحجوزات الهائلة. ويتمُّ تحديث مستودع بيانات سفريات المؤسسة بشكلٍ فوري تقريبًا بدفعات تعمل كل ١٥ دقيقة، ولجمع البيانات من جميع الأعمال التابعة لشركة Sabre. وتستخدم Sabre مستودع بيانات سفريات المؤسسة لإنشاء لوحات المعلومات التنفيذية التي توفر رؤى تنفيذية لحظية باستخدام منصة Cognos ذكاء الأعمال مع Oracle Data Integrator وOracle Goldengat. وتوفر لوحات المعلومات التنفيذية للمديرين وصُناع القرار رفيعي المستوى من شركات الطيران والعملاء حلًا آليًا سهل الاستخدام يجمع بين مقاييس الأداء الحساسة بطريقة موجزة وتقديم نظرة ثاقبة سريعة بدرجة ٣٦٠ عن صحة شركة الطيران. وفي إحدى شركات الطيران، توفر لوحات المعلومات التنفيذية لشركة Sabre للإدارة العليا صورةً يوميةً ولحظية لمؤشرات الأداء الرئيسة في تطبيقٍ واحد تحل محل العملية التي يتم إجراؤها مرةً واحدةً في الأسبوع والتي تستغرق ٨ ساعات لتوليد التقرير نفسه من

مصـادر البيانـات المختلفة. إن اسـتخدام اللوحات لا يقتصر عـلى العملاء الخارجيين بل تستخدمها Sabre أيضاً لتقييمها للأداء التشغيلي الداخلي.

وتسـاعد لوحات المعلومـات عملاء **Sabre** على الفهم الواضـح للبيانات من خلال شاشات العرض المرئية التي تتضمّن قدراتٍ تفاعلية، يستبدل العروض التقديمية المسطحة ويسـمح بمراجعة أكثر تركيزًا للبيانات بأقل جهد ووقت. هذا يُسـهِّل حوار الفريق من خلال إتاحة البيانات/ المقاييس المتعلقة بأداء المبيعات للعديد من المسـتفيدين، بما في ذلـك التذاكر، والمقاعد المباعة والطيران، والأداء التشـغيلي بما في ذلك البيانات المتعلقة بحركة الطيران والتتبُّع، وحجوزات العملاء، والمخازن، والإيرادات عبر توزيع شركة طيران متعددة القنوات. وتوفر أنظمة لوحة المعلومات البنية التحتية القابلة للتوسـع، ودعم واجهة المستخدم الرسومية، وتكامل البيانات، والتجميع الذي يُمكِّن المديرين التنفيذيين في شركات الطـيران من أن يكونوا أكثر إيجابيةً في اتخاذ إجراءات تؤدي إلى آثار إيجابية على الصحة العامة لشركات الطيران التابعة لهم.

وباسـتخدام ETDW الخـاص بها، تسـتطيع **Sabre** كذلك تطويـر حلول تحليلية وإعـداد تقارير أخرى تسـتندُ إلى الويب، والتي تتيح الحصول عـلى البيانات من أجل الحصـول على رؤى العمـلاء من خلال تحليل الملفات الشـخصية للعمـلاء وتفاعلات مبيعاتهم من أجل حساب قيمة العملاء. وهذا يتيح تقسيماً أفضل للعملاء وإحصاءات للخدمات ذات القيمة المضافة.

أسئلة للمناقشة:

١- ما هي التقارير التقليدية؟ وكيف يتمُّ استخدامها في المنظمة؟

٢- كيف يمكن استخدام التحليلات لتحويل التقارير التقليدية؟

٣- كيف يمكنُ للتقارير التفاعلية أن تساعد الشركات في صُنْع القرار؟

ما الذي الذي يمكننا أن نتعلمه من هذه الحالة العملية؟

تُوضِّح هذه الحالة العملية أن المنظمات التي استخدمت تقاريرها في وقتٍ سابق فقط لتتبـع أعمالها الداخلية وتلبية متطلبات التوافق التـي وضعتها الحكومة تتجه الآن نحو توليد معلومات ذكية قابلة للتطبيق من بيانات معاملاتها التجارية. وقد أصبحت التقاريرُ أكثر اتسـاعًا؛ إذ تحاول المنظمات الآن تحليل بيانات المعاملات المؤرشفة؛ لفهم التوجهات والأنماط الخفية الكامنة التي تمكِّنها من اتخاذ قرارات أفضل من خلال اكتساب رؤى حول

المجالات التي تنطوي على مشكلات وحلها لمتابعة فرص السوق الحالية والمستقبلية. وقد تطورت التقارير إلى تقارير تفاعلية عبر الإنترنت، وهو ما يمكّن المستخدمين من الحصول على بيانات سريعة وإنشائها، بمساعدة أدوات التصوير التي لديها القدرة على الاتصال بقاعدة البيانات؛ مما يوفر إمكانيات التعمق في البيانات الملخّصة.

Source: Teradata.com, "Sabre Airline Solutions," Terry, D. (2011), "Sabre Streamlines Decision Making," http://www.Teradatamagazine.com/v11n04/Features/Sabre-Streamlines-Decision-Making/(Accessed July 2016).

استعمال الوسائط المتعددة في ذكاء الأعمال:

تتضمّن شبكة جامعة تيراداتا TUN مقاطع الفيديو (على غِرار البرنامج التلفزيوني CSI)؛ لتوضيح مفاهيم التحليلات في مختلف الصناعات. ويُطلَق عليها اسم «مقاطع فيديو تقصّي سيناريوهات الأعمال BSI». وهي ليست لمجرد التسلية؛ ولكنها تزوّد الفصل الدراسي ببعض الأسئلة للمناقشة. وبالنسبة للمبتدئين، يُرجَى الانتقال إلى: http://www.teradatauniversitynetwork.com/ Library/Items/BSI-The-Case-of-the-Misconnecting-Passengers أو www.youtube.com/ watch?v=NXEL5F4_aKA ومشاهدة الفيديو الذي يظهر على يوتيوب. وعليك أن تفترض قيامك بدور مركز خدمة العملاء المهنية. وقد بدأت الرحلة القادمة في وقتٍ متأخر، ومن المرجَّح أن يفقد العديد من الركاب رحلاتهم. وهناك مقاعدُ في رحلة مغادرة واحدة يمكن أن تستوعب اثنين من الركاب الأربعة. فأيٌّ من الركاب يجب أن تُعطَى إليه الأولوية؟ ويتمُّ إعطاؤك معلومات حول ملامح العملاء والعلاقات بشركة الطيران. وهنا قد تتغيرُ قراراتك عندما تتعلم المزيد عن ملفات التعريف لهؤلاء العملاء.

شاهد الفيديو وأوقفه كلما لزم الأمر، وأجب عن الأسئلة التي ينبغي أن يُعطَى فيها الأولوية للمستفيدين. ثم استأنِف مشاهدة الفيديو؛ للحصول على مزيدٍ من المعلومات. وبعد اكتمال الفيديو، يمكنك مشاهدة الشرائح المتعلقة بهذا الفيديو وكيف تمَّ إعداد التحليل على شريحة مُحدَّدة في:

www.slideshare.net/teradata/bsi-how-we-did-it-the-case-of-the-misconnecting-passengers.

وتوفِّر هذه الرحلة المتعددة الوسائط مثالاً على كيفية مساعدة المعلومات الإضافية المتاحة من خلال مشروع مستودع البيانات DW في صُنع القرار.

على الرغم من أن البعض يساوي بين نظام دعم القرار وذكاء الأعمال؛ فإنَّ هذه الأنظمة ليست في الوقت الحاضر متساويةً. المثير للاهتمام أن نلاحظ أن بعض الناس يعتقدون أن نظام دعم القرار DSS هو جزءٌ من ذكاء الأعمال وإحدى أدواتها التحليلية. ويعتقد آخرون أنَّ ذكاء

الأعــمال هو حالة خاصة من أنظمة دعم القــرار التي تتناول في الغالب إعداد التقارير والتواصل والتعاون (شــكل من أشــكال أنظمة دعم القــرار الموجهة بالبيانات). وثمَّة تفســير آخر قال به (Watson، 2005) هــو أن ذكاء الأعــمال ما هو إلا نتيجة لتطوُّر مســتمر، وعلى هذا النحو؛ فإن نظــام دعم القرار هــو أحد عناصر ذكاء الأعــمال الأصلية. وعلاوةً على ذلــك، وكما هو مُبيَّن في القسم التالي وما بعده؛ نجد أنه وفي العديد من الدوائر، تم دمجُ ذكاء الأعمال من خلال تحليلات مصطلحات جديدة أو علم البيانات.

معالجة العمليات مقابل المعالجة التحليلية:

لتوضيح الخصائص الرئيســة لذكاء الأعمال؛ سنقوم أولًا ببيان أن ذكاء الأعمال لا يُعَدُّ معالجةً للمعامــلات. إننا جميعًا على دراية بأنظمة المعلومات التي تدعم معاملاتنا، مثل أجهزة الصراف الآلي، والودائــع المصرفية، وفحص النقود في متجر البقالة، وما شــابهها. وتشــارك أنظمة معالجة العمليــات هــذه باســتمرار في التعامل مــع التحديثات فيما يمكن أن نســميه قواعــد البيانات التشــغيلية. فعلى ســبيل المثال: في عملية الســحب من أجهزة الــصراف الآلي، نكون بحاجةٍ إلى تقليل رصيدنا المصرفي وفقاً لذلك؛ في حين أنَّ الإيداع البنكي يضيف إلى الحساب. ومن المرجَّح أن تظهر مشتريات متجر البقالة في حساب المتجر من خلال إجمالي المبيعات لهذا اليوم، وينبغي أن تعكس تخفيضاً مناســباً في المخزون من الأصناف التي اشــتريناها وهكذا. وتعالج أنظمة معالجة العمليات الفورية هذه (OLTP) نشــاط الشركة المعتاد والمســتمر. وعلى النقيض من ذلك؛ فإن مستودع البيانات يكون عادةً نظامًا متميزًا يوفِّر التخزين للبيانات التي سيتم استخدامها للتحليل. والقصد من هذا التحليل؛ هو إعطاء الإدارة القدرة على تنظيف البيانات للحصول على معلومات عــن الأعمال، ويمكن اســتخدامها لتوفير الدعم التكتيكي أو التنفيذي للقــرار؛ إذ يُمكِن للمديرين المباشرين، على ســبيل المثال، اتخاذ قرارات أسرع و/أو أكثر اســتنارة. وســوف نقدِّم تعريفًا أكثر تقنيةً لمســتودع البيانات في الفصل الثاني، ولكن يكفي أن نقول بأن مســتودعات البيانات تهدف إلى العمــل مع البيانات المعلوماتية المُســتَخدَمة في نظم المعالجــة التحليلية الفورية (OLAP). ويتمُّ تخزين معظم البيانات التشــغيلية في نظم تخطيط موارد المؤسســة (ERP) وفي أشــقائها التكميليين، مثل إدارة سلســلة التوريد (SCM) أو نظــام إدارة العملاء (CRM) المخزَّنة في نظام معالجة العمليات الفورية OLTP، وهو نوعٌ من معالجة الحاســوب حيث يســتجيب الحاسوب فورًا لطلبات المســتخدمين. ويُعَدُّ كلُّ طلب معاملةً، وهو ســجلٌ محوسب لحدثٍ منفصل، مثل اســتلام المخزون أوامر العميل. وبعبــارةٍ أخرى؛ تطلب المعاملة مجموعــةً من اثنين أو أكثر من تحديثات قاعدة البيانات التي يجب أن تكتمل بطريقة كلِّ شيء أو لا شيء.

إن التصميمَ الذي يجعل نظام OLTP هذا فعَّالًا لمعالجة العمليات يجعله غير فعال بالنسبة للتقارير والاستعلامات والتحليلات الخاصة بالمستخدِم النهائي. وفي الثمانينيات، أشار العديدُ من المستخدمين من رجال الأعمال إلى إطاراتهم الرئيسة على أنها «ثقوب سوداء» لأن كلَّ المعلومات دخلت إليها، ولكن لم يرجع أيٌّ منهم على الإطلاق. ويجب أن يبرمج موظفو تقنية المعلومات جميع طلبات تقديم التقارير، في حين أنه يمكن إصدار التقارير «المسبقة أو الأكثر استخداماً» وحدها على أساس مجدول، إن الاستعلام الفوري المخصص مستحيل تقريبًا. وعلى الرغم من أن نظم تخطيط موارد المؤسسات القائمة على العميل/ الخادم في التسعينيات كانت أكثر ملاءمةً لإعداد التقارير؛ فإنها ما زالت بعيدة كلَّ البعد عن الاستخدام المرغوب فيه من قِبل المستخدمين النهائيين وغير التقنيين لأمور مثل إعداد التقارير التشغيلية والتحليل التفاعلي وما إلى ذلك. ولحلِّ هذه القضايا؛ تمَّ إنشاء مفاهيم مستودع البيانات وذكاء الأعمال.

وتحتوي مستودعات البيانات على مجموعةٍ واسعةٍ من البيانات التي تقدِّم صورةً متماسكةً لأوضاع الأعمال في وقتٍ واحد. وكانت الفكرة هي إنشاء بنية تحتية لقاعدة البيانات التي كانت دائمًا على شبكة الإنترنت وتتضمَّن جميع المعلومات من أنظمة OLTP، بما في ذلك البيانات التاريخية، ولكن أُعِيد ترتيبها وتنظيمها بطريقة سريعة وفعَّالة للاستفسار والتحليل ودعم القرار. إنَّ فصل OLTP عن التحليل ودعم القرار تمكِّن من الاستفادة من ذكاء الأعمال وهو ما وصفناه في وقتٍ سابق.

التخطيط المناسب والتوافق مع إستراتيجية الأعمال:

أولاً وقبل كلِّ شيء، يجب أن تكون الأسباب الأساسية للاستثمار في ذكاء الأعمال متوافقةً مع إستراتيجية أعمال الشركة. إنَّ ذكاء الأعمال ليس مجرد ممارسة تقنية لقسم نظم المعلومات؛ بل يجب أن يكون وسيلةً لتغيير الطريقة التي تقوم بها الشركة التجارية من خلال تحسين عملياتها التجارية وتحويل عمليات صُنع القرار؛ لتكون قائمةً إلى حَدٍّ كبير على البيانات. ويشير العديدُ من استشاريي وممارسي ذكاء الأعمال الذين يشاركون في مبادرات ذكاء الأعمال الناجحة إلى أن إطار التخطيط هو شرطٌ مسبقٌ ضروري. يقوم أحد إطارات العمل الذي تم تطويره بواسطة مجموعة جارتنر (2004) Gartner بتحليل وتخطيط وتنفيذ مكونات الأعمال والمنظمة والوظائف والبنية التحتية. وعلى مستوى الأعمال والمستويات التنظيمية؛ يجب تحديدُ الأهداف الإستراتيجية والتشغيلية مع مراعاة المهارات التنظيمية المتاحة لتحقيق تلك الأهداف. ويجب على الإدارة العليا أن تنظر في قضايا الثقافة التنظيمية التي تحيط بمبادرات ذكاء الأعمال وحفز تلك المبادرات والإجراءات المتعلقة بمشاركة أفضل ممارسات ذكاء الأعمال المتبعة في الإدارة العليا داخل الشركة، مع وَضْع الخطط اللازمة لإعداد المنظمة من أجل التغيير. ومن بين الخطوات الأولى في هذه

العملية تقييم تنظيم نظام المعلومات ومجموعات مهارات الفئات المحتملة من المستخدمين، وما إذا كانت الثقافة قابلةً للتغيير. ومن خلال هذا التقييم، وعلى افتراض وجود مبرر/ ضرورة للمُضِي قُدُمًا، يمكن للشركة إعداد خطة عمل مفصلة. وثمَّة مسألة حاسمة أخرى لنجاح تنفيذ نظم المعلومات وهي دَمْج العديد من مشاريع ذكاء الأعمال (حيث تستخدم معظم المؤسسات العديد من مشاريع ذكاء الأعمال) فيما بينها من جانب ومع أنظمة تقنية المعلومات الأخرى في المنظمة وشركائها التجاريين من جانبٍ آخر.

من الحكمة البدءُ في استخدام ذكاء الأعمال وإنشاء مركز كفاءة ذكاء الأعمال إذا كانت إستراتيجية الشركة تتماشى بشكلٍ صحيح مع أسباب مبادرات مستودعات البيانات وذكاء الأعمال، وإذا كان نظام المعلومات قادرًا أو بإمكانه القيام بدوره في مثل هذا المشروع، وإذا كان مجتمع المستخدمين المستهدَف موجوداً ولديه الدافعية؛ حينها بإمكان المركز أن يقدم الخدمات التالية (Gartner، 2004):

- يمكِن للمركز أن يُوضِّح كيف يرتبط ذكاء الأعمال بالإستراتيجية وتنفيذ الإستراتيجية.
- يمكِن أن يعمل المركزُ على تشجيع التفاعل بين مجتمعات مستخدمي الأعمال المحتملين وتنظيم نظم المعلومات.
- يمكِن للمركز أن يعمل كمستودع ومُوزِّع لأفضل ممارسات ذكاء الأعمال بين مختلف خطوط الأعمال.
- يمكِن للمركز تأسيسُ معايير للتميُّز في ممارسات ذكاء الأعمال وتشجيعها من خلال الشركة.
- يمكِن أن تتعلَّم منظمات نظم المعلومات الكثير من خلال التفاعل مع مجتمعات المستخدمين، مثل المعرفة حول مجموعة متنوعة من الأدوات التحليلية المطلوبة.
- يمكِّن مستخدمي الأعمال التجارية ومنظمات نظم المعلومات من فهمٍ أفضل للسبب الذي يستوجب أن تكون مِنصة مستودع البيانات مرنةً بما فيه الكفاية لتوفير متطلبات العمل المتغيرة.
- يمكِن أن تساعد حاملي الأسهم المهمين، مثل المديرين التنفيذيين رفيعي المستوى؛ على فهم كيف يمكِن أن يلعب ذكاءُ الأعمال دورًا مهمًّا.

ومن عناصر ذكاء الأعمال الأخرى الناجحة قدرته على توفير بيئة فاعلة سنعرضها فيما يلي:

ذكاء الأعمال الفوري وعند الطلب أصبح متاحاً:

لقد تزايدت الحاجةُ إلى الوصول الفوري إلى المعلومات المشتتة بناءً على الطلب من أجل سد الفجوة بين البيانات التشغيلية والأهداف الإستراتيجية، وهو الأمر الذي أصبح أكثر إلحاحًا. ونتيجةً لذلك؛ ظهرت فئةٌ من المنتجات تُسمَّى تطبيقات ذكاء الأعمال الفورية. ومن شأن إدخال تقنيات جديدة لتوليد البيانات، مثل تقنية RFID وغيرها من أجهزة الاستشعار، أن يؤدي إلى

تسريع هذا النموِّ وتلبية الحاجة إلى ذكاء الأعمال بشــكلٍ فوري. وتســتخدم أنظمة ذكاء الأعمال التقليديــة قــدرًا كبيرًا من البيانــات الثابتة التي تمَّ اســتخراجها وتنقيتها وتنزيلها إلى مســتودع البيانــات لإنتاج التقارير والتحليلات. ومع ذلك؛ فإن الحاجة ليســت مجــرد إعداد التقرير؛ لأن المستخدمين يحتاجون إلى مراقبة الأعمال، وتحليل الأداء، وفهم لماذا تحدث الأمور. وهذا يمكِن أن يساعِد المستخدمين، الذين يحتاجون أن يعرفوا (تقريبًا بشكل فوري) معلومات عن التغييرات في البيانات أو توفُّر التقارير ذات الصلة، والتنبيهات، والإخطار بشــأن الأحداث والاتجاهات الناشئة في تطبيقــات وســائل التواصل الاجتماعــي. وبالإضافة إلى ذلك؛ يمكِن برمجــة تطبيقات الأعمال للعمل على ما تكتشفه أنظمة ذكاء الأعمال هذه في الوقت المناسب. على سبيل المثال: قد يقوم (SCM) بتقديــم طلبٍ تلقائي؛ للحصــول على المزيد من «الأدوات» عندمــا يكون المخزون أقلَّ مــن حَدٍّ معين أو حينما يقوم نظام خدمة العملاء بتنبيــه ممثل خدمة العملاء وموظف مراقبة الائتمان للتحقق من عملية عبر الإنترنت قيمتها تجاوزت ١٠٠٠٠ دولار.

من الطرق التي يســتخدمها ذكاء الأعمال لتوفير البيانات في الوقت المناسب نموذج مستودع البيانات الذي تســتخدمه أنظمة ذكاء الأعــمال التقليدية. وفي هذه الحالة؛ توفِّر المنتجات التي يقدِّمهــا مــورِّدو منصة ذكاء الأعمال المبتكرة حلًا موجهًا بالخدمة وبشــكل فوري تقريبًا، والذي يقوم بتعبئة مستودع البيانات (DW) بشكلٍ أسرع مما يقوم به التحديث الدوري التقليدي من اســتخراج/ نقل/ تنزيل (انظر الفصل ٣). الطريقة الثانية تُسمَّى بإدارة نشاط الأعمال (BAM) فيتــمُّ اســتخدامه من قِبل مورِّدي BAM و/أو مورِّدي وســطاء BAM معًــا (مثل: Savvion أو برنامج Iteration أو Vitria أو webMethods أو Quantive أو Tibco أو Vineyard). وهو يتجاوز مســتودع البيانات تمامًا ويســتخدم خدمات الويب أو غيرها من وسائل الرَّصد لاكتشاف الأحــداث التجارية الرئيســة. ويمكن وَضْع أجهزة مراقبة البرنامــج (أو وكلاء الذكاء) على خادم منفصل في الشــبكة أو في قواعد بيانات تطبيقات العمليات نفسها، ويمكنهم استخدام الأساليب القائمة على العمليات؛ من أجل قياس العمليات التشغيلية بشكلٍ استباقي وبذكاء.

تطوير أو حيازة أنظمة ذكاء الأعمال:

يقــدِّم اليوم العديــدُ من المورِّديــن أدواتٍ متنوعةً البعض منها مبرمجٌ بشــكلٍ كاملٍ ويُطلَق عليها Shells، وكلُّ ما يجبُ على المستخدِم فعله هو إدخال الأرقام الخاصة به، ويمكن شراء هذه الأدوات أو تأجيرها. وللحصول على قائمة بالمنتجات والعروض التوضيحية والمزيد من المعلومات عــن المنتجــات الحالية، انظر أدلة المنتج على موقع معهد مســتودعات البيانات tdwi.org. وما عليك إلا أن تقوم بتسجيل المستخدِم مجانًا. ويتمُّ إنشاء جميع تطبيقات ذكاء الأعمال تقريبًا مع

shells والتي يوفرها الموزِّدون أنفســهم الذين يقومون بإنشاء حلٍّ مخصص للعميل أو العمل مع مورِّد آخر يستعين بمصادر خارجية. إنَّ المشكلة التي تواجهها الشركات هي اختيار البديل: الشراء أو الإيجــار أو الإنشــاء. ولكلٍّ من هــذه البدائل عِدة خيارات، ومن بين المعايير الرئيســة لاتخاذ القرار؛ هو التحليل وتحليل التكلفة والمنافع.

الضبط وتحليل التكاليف والمنافع:

مع زيادة عدد تطبيقات ذكاء الأعمال المحتملة؛ تنشــأ الحاجة إلى ضبطها وتحديد أولوياتها. وهذه ليســت مهمة سهلة بســبب العدد الكبير من الفوائد غير المباشرة؛ إذ يجبُ تحديد الفوائد المبــاشرة وغير المباشرة. وبطبيعة الحــال؛ فإنَّ ذلك يحدث عندما تكون معرفة تطبيقات مختلفة في المنظمات الأخرى ودراســات الحالة مفيدة للغاية. على ســبيل المثال: يوفر معهد مستودعات البيانات (tdwi.org) ثروة من المعلومات عن المنتجات والتطبيقات والتطبيقات المبتكرة. ويمكن أن تكون هذه المعلومات مفيدةً في تقدير الفوائد المباشرة وغير المباشرة.

أمن وحماية الخصوصية:

هذه مسألةٌ مهمةٌ للغاية في تطوير أيِّ نظام محوسب، وخاصةً ذكاء الأعمال الذي يحتوي على البيانات التي قد يكون لها قيمة إستراتيجية. وكذلك يجبُ حماية خصوصية الموظفين والعملاء.

تكامل الأنظمة والتطبيقات:

باســتثناء بعض التطبيقــات الصغيرة؛ يجب تكامل جميع تطبيقــات ذكاء الأعمال مع أنظمة أخرى، مثل: قواعد البيانات، والأنظمة القديمة، ونظم المؤسســة (خاصةً ERP وCRM) والتجارة الإلكترونيــة (البيع والشراء)؛ بل وأكثر من ذلــك. وبالإضافة إلى ذلك؛ فإن تطبيقات ذكاء الأعمال ترتبط عادةً بالإنترنت وفي أحيان كثيرة بنظم معلومات شركاء الأعمال.

وعــلاوةً عــلى ذلك؛ تحتــاج أدوات ذكاء الأعمال أحيانًا إلى التكامل والدعــم فيما بينها. وقد دفعت الحاجة إلى التكامل مورِّدي البرامج إلى إضافة قدراتهم باستمرار إلى منتجاتهم. إن العملاء الذين يشترون حزمة البرمجيات المتكاملة يتعاملون مع مورِّدٍ واحدٍ فقط ولا يحتاجون إلى التعامل مع أنظمة أخرى. ولكنهم قد يفقدون مَيْزة إنشاء نظم تتألف من مكوِّنات «من سلالة أفضل».

أسئلة مراجعة على القسم ١-٤:

١- عرِّف ذكاء الأعمال.

٢- اذكر مع الشرح المكونات الرئيسة لذكاء الأعمال.

٣- عرِّف OLTP.

٤- عرِّف OLAP.

٥- اذكر بعض مجالات التنفيذ التي تناولها تقرير Gartner.

٦- اذكر بعض عوامل النجاح الأخرى لذكاء الأعمال.

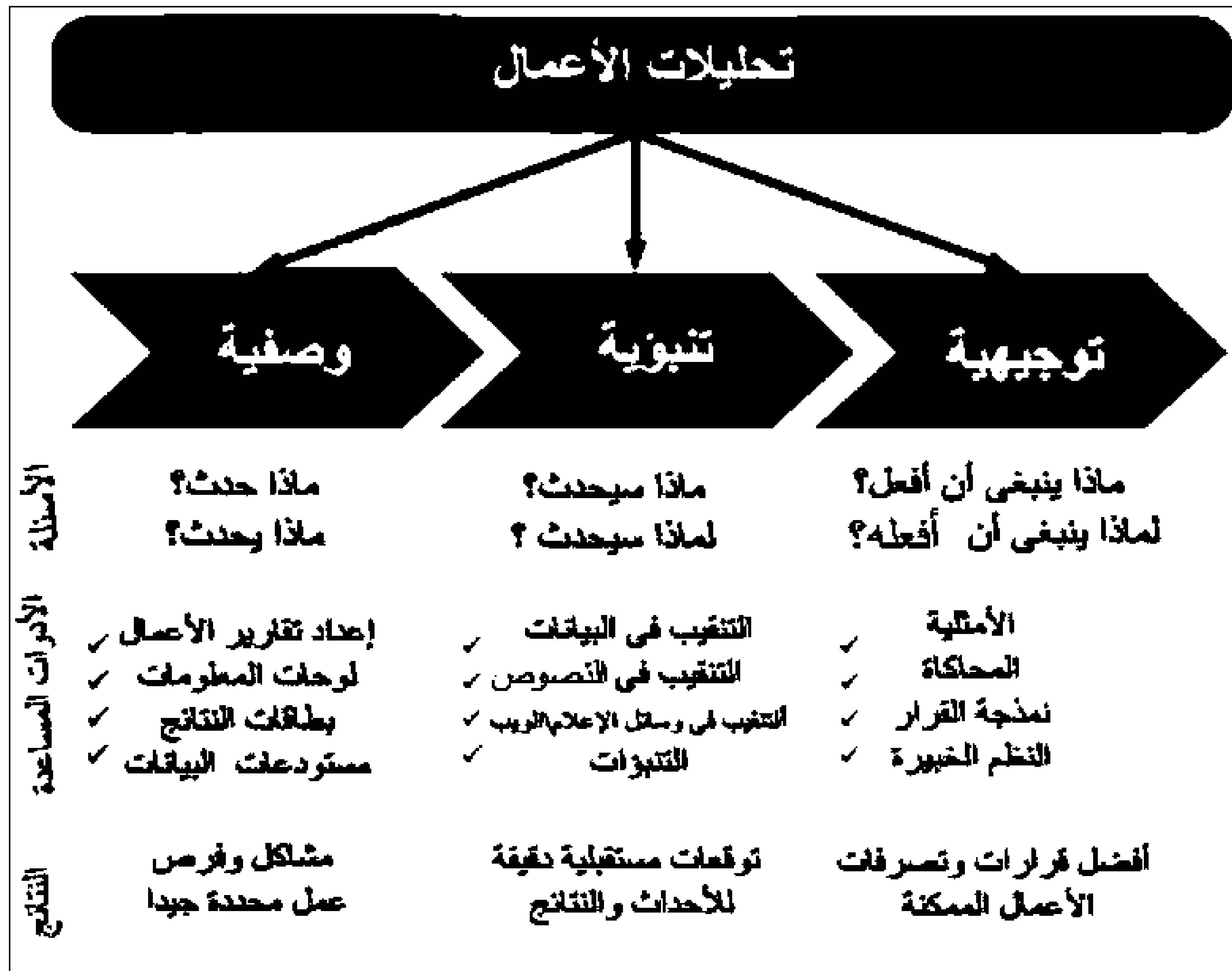
١-٥ نظرة عامة على التحليلات:

لقــد حلَّت كلمــة التحليلات إلى حَدٍّ كبير محل المكوِّنات الفردية الســابقة لتقنيات دعم القرار المحوســبة التــي كانت متوفرةً تحت تســميات مختلفة في المــاضي. وفي الواقع؛ فــإن العديد من الممارسين والأكاديميين الآن يستخدمون كلمة التحليلات بدلًا من ذكاء الأعمال. وعلى الرغم من أن العديد من المؤلفين والاستشــاريين عرَّفوها بشــكلٍ مختلف قليلاً؛ فقد يُنظر إلى التحليلات على أنها عملية وَضْع قرارات قابلة للتطبيق أو توصيات لاتخاذ إجراءات بناءً على رؤى متولدة من البيانات التاريخية. ووفقاً لمعهد بحوث العمليات وعلوم الإدارة (INFORMS)؛ فإنَّ التحليلات تمثل مزيجاً من تقنيات الحاسوب، وتقنيات علم الإدارة، والإحصاءات لحلِّ مشكلات حقيقية. وبطبيعة الحال؛ فقد اقترحت منظمات أخرى كثيرة تفســيراتها ودوافعها للتحليلات. فعلى سبيل المثال: اقترح معهد SAS ثمانية مستويات من التحليلات تبدأ بتقارير موحَّدة من النظم الحاسوبية وتوفر هذه التقارير وصفاً لما يحدث داخل المنظمة، وقد مكَّنتنا التقنيات الإضافية من إنشاء تقارير مُخصَّصة بتفاصيل أكثر. ويعود التطوُّر التالي في إعداد التقارير إلى طلبات البحث من نوع المعالجة التحليلية الفورية التي تسمح للمستخدِم بالاستكشاف بشكلٍ أعمق وتحديد مصادر معينة للمشكلات أو الفرص.

التقنيات المتاحة اليوم لديها القدرة على إصدار تنبيهات لصناع القرار حينما يسوء الأداء على ســبيل المثال، وقد تكون هذه التنبيهات للمســتهلك فيما يتعلق بالطقس أو قضايا أخرى. ولكن يمكن على غِرار ذلك إنشاء التنبيهات في سياق إعدادات معينة عندما تكون المبيعات أعلى أو أقل من مســتوى معين خلال فــترة زمنية معينة أو عندما يكون المخــزون لمنتج معين منخفضاً. وكلُّ هذه التطبيقات تكون ممكنةً من خلال تحليل البيانات التي يتمُّ جمعها من قِبل منظمةٍ ما، وقد يســتلزمُ هذا التحليل تحليلاً إحصائياً لفهم الأنماط بشــكلٍ أفضل. ويمكن بعد ذلك اتخاذ خطوة أخرى لتطوير التنبؤات أو نماذج للتنبؤ بكيفية استجابة العملاء لحملة تسويقية معينة أو عروض خدمات/ منتجات مســتمرة. وعندما تكون لــدى المنظمة رؤية جيدة لما يحدث وما قد يحدث، يمكنها أيضاً أن تســتخدم أساليب أخرى لإصدار أفضل القرارات في ظل الظروف القائمة. وتُصنَّف

على ثمانية مستويات للتحليلات ولمزيدٍ من التفاصيل بالإمكان النظر في العرض التوضيحي المقدَّم من معهد SAS على الرابط: (sas.com/news/sascom/analytics_levels.pdf).

ولقد جسَّدَت هذه الفكرة المتمثلة في دراسة البيانات بشكلٍ كامل لفهم ما يحدث، وما سيحدث، وكيفية تحقيق أفضل استفادة منها في اقتراح ثلاثة مستويات من التحليلات. وتمَّ تحديدُ هذه المستويات الثلاثة (المعلوماتية/ المجتمعية/ التحليلية) على أنها وصفية، وتنبؤية، وتوجيهية. يعرض الشكل ١-١١ عرضاً بيانياً لأنواع التحليل الثلاثة، ويشير إلى أن هذه المستويات الثلاثة مستقلةٌ إلى حَدٍّ ما، وأن كلاً منها يؤدي إلى الآخر. كما يشير إلى وجودِ بعض التداخل عبر هذه الأنواع الثلاثة من التحليلات. وسواءً كانت مستقلةً أو متداخلةً؛ فإن الطبيعة المترابطة لمختلف أنواع التحليلات واضحةٌ، وسيتم شرح المستويات الثلاثة في الأقسام التالية.



الشكل ١-١١: أنواع التحليلات الثلاثة

التحليلات الوصفية:

تشيرُ التحليلات الوصفية (أو التقارير) إلى معرفة ما يحدث في المنظمة وفهم بعض الاتجاهات والأسباب الكامنة وراء مثل هذه التجارب. ينطوي ذلك على توحيد مصادر البيانات وتوفر

جميـع البيانات ذات الصلة في شـكلٍ يتيحُ إعداد التقارير والتحليلات المناسـبة، عادةً ما يكون تطوير هذه البنية الأساسـية للبيانات جزءًا من مستودعات البيانات. ومن هذه البنية الأساسية للبيانات يمكننا تطويرُ التقارير المناسبة، والاستعلامات، والتنبيهات، والاتجاهات باستخدام أدوات وأساليب إعداد التقارير المتنوعة.

من التقنيات المهمة التي تشكُّل لاعباً أساسيًّا في مجال التحليلات الوصفية؛ التصويرُ باستخدام أحدث أدوات التصوير في السـوق، والتي تمكِّننا من اسـتخراج رؤى قوية من عمليات منظماتنا. (dssbibook.com).

حالة عملية ١-٢

شركة سيلفاريس (Silvaris) توسِّع أعمالها باستخدام التحليل المرئي، وإمكانيات إعداد التقارير الفورية

تأسَّسـت شركة Silvaris في عام ٢٠٠٠م من قِبل فريـقٍ من المتخصِّصين في صناعة أخشـاب الغابات لتوفير التقدُّم التقني في قطاع الخشـب ومواد البناء. وتُعَدُّ Silvaris أولَ منصـةٍ للتجارة الإلكترونية في الولايات المتحدة الأمريكية خصيصاً لمنتجات الغابات، ويقع مقرُّها الرئيس في سـياتل، واشـنطن. وهي شركة رائدة في مجال توفير المنتجات الخشبية الصناعية وفائض مواد البناء.

تقـوم Silvaris ببَيْـع منتجاتها وتُقدِّم خدمات لوجيسـتية دوليـة إلى أكثر من ٣٥٠٠ عميل. ولإدارة مختلف العمليات ذات الصِّلة بالمعاملات؛ أنشأت منصة تداول عبر الإنترنت بحقوق مِلكية لتتبع تدفق المعلومات المتعلِّقة بالمعاملات بين التجار والمحاسـبة والائتمان والخدمات اللوجسـتية. وقد أتاح ذلك للشركة تبادل المعلومات بشـكلٍ فوري مع عملائها والشركاء. ولكن نظراً للتغيُّر السريع في أسعار المواد؛ أصبح من الضروري أن تحصل Silvaris على عرضٍ فوري للبيانات دون الحاجة إلى نقل وتنسيق البيانات في تقرير منفصل.

وبـدأت Silvaris بالاسـتعانة بشركة تابلوه (Tableau)؛ لقدرتهـا على التواصل مع البيانات الحية وتصويرها. وبسـبب لوحات المعلومات التي أنشأتها تابلوه (Tableau) والتي تتسـم بسـهولة الفهم والـشرح؛ بدأت Silvaris باسـتخدام تابلـوه (Tableau) لأغراض إعداد التقارير. وسـاعد ذلك Silvaris في سحب المعلومات بسرعة من البيانات وتحديد القضايا التي تؤثر على أعمالها. ونجحت Silvaris في إدارة الطلبات عبر الإنترنت مقابل الطلبات الأخرى بمسـاعدة التقارير التي أنشأتها تابلوه (Tableau). والآن؛ تقوم Silvaris بتتبُّع الطلبات عبر الإنترنت المقدمة من قِبل العملاء وتعرف متى ترسل دفعات

تجديــد للعملاء الذين يقومون بالشراء عبر الإنترنت. وأيضاً يمكن لمحللي Silvaris توفير الوقــت من خلال توليد لوحات معلومات بدلاً من كتابــة مئات الصفحات من التقارير باستخدام تابلوه (Tableau).

أسئلة للمناقشة:

١- ما التحدي الذي واجهته Silvaris؟

٢- كيف حلّت Silvaris مشكلتها باستخدام تصوير البيانات مع تابلوه (Tableau)؟

ماذا يمكننا أن نتعلمَ من هذه الحالة العملية:

تحتــاج العديــدُ من الصناعات إلى تحليل البيانات بشــكلٍ فــوري. ويتيح التحليلُ الفوري للمحللين تحديد القضايا التي تؤثر على أعمالهم. ويكون التصويرُ أحيانًا أفضل طريقــة لبــدء تحليل تيارات البيانات الحيــة. وتُعَدُّ تابلــوه (Tableau) إحدى أدوات تصويــر البيانــات ذات القدرة على تحليل البيانات الحيــة دون نقل البيانات الحية إلى تنسيق تقارير منفصل.

Sources: Tableau.com, "Silvaris Augments Proprietary Technology Platform with Tableau's Real-Time Reporting Capabilities," http://www.tableau.com/sites/default/files/case-studies/silvarisbusiness-dashboards_0.pdf (accessed July 2016); Silvaris.com, "Overview," http://www.silvaris.com/About/(accessed July 2016).

حالة عملية ١-٣

شركة سيمنس (Siemens) تقلِّل التكلفة باستخدام تصوير البيانات

شركــة ســيمنس (Siemens) هي شركة ألمانية مقرُّها في برلــين، ألمانيا، وهي إحدى كــبرى الــشركات في العالم التــي تركِّز على مجــالات الكهربــاء، والســيارات، والرقمنة. وتحقق إيرادات سنوية تبلغ ٧٦ مليار يورو.

ويتــمُّ تنفيذ مجموعــة التحليلات المرئية لشركة ســيمنس (Siemens) بحلــول التقارير الشــاملة والاستشــارات فيما يتعلق بجميع احتياجات ذكاء الأعمال الداخلية لشركة سيمنس (Siemens). وكانــت هذه المجموعة تواجه تَحَدِّياً متمثــلاً في توفير حلول إعداد التقارير إلى منظمة سيمنس (Siemens) بالكامل عبر الأقسام المختلفة مع الحفاظ على التوازن بين قدرات الحوكمــة والخدمة الذاتية، وكانت شركة ســيمنس (Siemens) بحاجــةٍ إلى مِنصة تُمكِّنها من تحليل الحالات المتعددة لاســتطلاعات آراء العملاء، والعمليات اللوجيســتية، وإعداد التقارير

المالية. وكان لا بد أن تكون هذه المنصة سهلة الاستخدام لموظفيها؛ حتى يتمكّنوا من استخدام هذه البيانات للتحليل واتخاذ القرارات. وبالإضافة إلى ذلك؛ يجبُ أن تتكامل المنصة بسـهولةٍ مع أنظمة سيمنس (Siemens) الحالية، وأن تمنح الموظفين تجربة استخدام سلسة.

بدأت الشركة في الاسـتعانة بشركة Dundas BI، وهي موردٌ عالمي رائدٌ لحلول ذكاء الأعمال وحلول التمثيل البصري للبيانات، وقد سـمح ذلك لشركة سـيمنس (Siemens) بإنشـاء لوحـات معلومـات تفاعلية مكّنَت شركة سـيمنس (Siemens) من اكتشـاف المشـكلات مبكرًا، وبالتالي توفير قدرٍ كبيرٍ من المال. وسـاعدت لوحات المعلومات التي طوّرتها Dundas BI شركة سـيمنس (Siemens) على الإجابة عن أسئلة مثل كيف تؤثر معدلات العرض المختلفة في المواقع المختلفة على العملية؛ مما سـاعدها على تقليل زمن دورة الأعمال بمقدار ١٢٪ وتكلفة الخردة بنسبة ٢٥٪.

أسئلة للمناقشة:

١- ما التحديات التي واجهتها مجموعة سيمنس (Siemens) للتحليلات المرئية؟

٢- كيف ساعدت أداة تصوير البيانات من Dundas BI شركة سيمنس (Siemens) في خفض التكلفة؟

ماذا يُمكِننا أن نتعلم من هذه الحالة العملية:

تحتاج العديدُ من المنظمات إلى أدوات تمكّنها من تحليل بياناتها المتوفرة في أقسـام متعددة، هذه الأدوات تسـاعد في تحسـين أداء عمل المنظمات، وجَعْل عملية اكتشاف البيانات شفافةً بالنسبة للمستخدمين؛ لتمكينهم من تحديد مشكلات الأعمال بسهولة.

Sources: Dundas.com, "How Siemens Drastically Reduced Cost with Managed BI Applications," http://www.dundas.com/resource/getcasestudy?caseStudyName=09-03-2016-Siemens%2FDundas-BI-Siemens-Case-Study.pdf (accessed July 2016); Wikipedia.org, "SIEMENS," https://en.wikipedia.org/wiki/Siemens (accessed July 2016); Siemens.com, "About Siemens," http://www.siemens.com/about/en/(accessed July 2016).

التحليلات التنبؤية:

تهدف التحليلات التنبؤية إلى تحديدِ ما يمكن أن يحدث في المستقبل، وتعتمد هذه التحليلات على الأسـاليب الإحصائية بالإضافة إلى الأسـاليب الأخرى التي تمَّ تطويرهـا حديثًا والتي تندرج تحـت الفئـة العامة التنقيب في البيانـات. وتهدف هذه التقنيات إلى القـدرة على التنبؤ بما إذا

كان من المرجح أن ينتقل العميلُ إلى المنافسين، وماذا يجب على العميل شراؤه بعد ذلك وبأيِّ مقدار، وما هي العروض الترويجية التي قد يستجيبُ لها العميل، وما إذا كان هذا العميل جديراً بالثقة الائتمانية أم لا، وما إلى ذلك. العديدُ من التقنيات تُستخدَم في تطوير تطبيقات التحليلات التنبؤية بما في ذلك خوارزميات التصنيف المختلفة. فعلى سبيل المثال، وكما سنرى في الفصلين الرابع والخامس، يمكننا استخدام تقنيات التصنيف مثل الانحدار اللوجستي، ونماذج شجرة القرارات، والشبكات العصبية للتنبؤ بمدى جودة الصور المتحركة أمام شباك التذاكر. ويمكننا أيضاً استخدامُ خوارزميات التجميع لتقسيم العملاء إلى مجموعات مختلفة؛ حتى يكون بالإمكان استهدافهم بعروض ترويجية مُحدَّدة لهم. وأخيراً؛ يمكننا استخدامُ أساليب الترابط في المجتمعات لتقدير العلاقات بين السلوكيات الشرائية المختلفة، أيْ: إذا اشترى أحد العملاء منتجًا ما؛ فما الذي قد يشتريه العميل أيضاً؟ يمكن أن يساعد هذا التحليل بائع التجزئة في التوصية أو الترويج للمنتجات ذات الصلة. على سبيل المثال؛ فإن أيَّ بحثٍ عن منتج على Amazon.com ينتج عنه أن يقترح بائعُ التجزئة أيضاً منتجات مشابهة أخرى قد يهتم بها العميل. وسنقوم بدراسة هذه الأساليب وتطبيقاتها في الفصول من الثالث إلى السادس. وتوضِّح الحالة العملية ١-٤ إحدى هذه التطبيقات في الألعاب الرياضية.

حالة عملية ١-٤

تحليل الإصابات الرياضية

إن أيَّ نشاطٍ رياضي يكون عُرضةً لوقوع إصابات وإذا لم يتم التعامل مع الإصابات بشكلٍ صحيح؛ فقد يتضرَّر الفريق. ويمكِن أن يساعد استخدام التحليلات لفهم الإصابات في الحصول على إحصاءات قيِّمة من شأنها تمكين المدربين وأطباء الفريق من إدارة تكوين الفريق، وفهم قدرات اللاعبين، والمساعدة في نهاية المطاف في اتخاذ القرارات الأفضل بشأن اللاعبين الذين ربما يكونوا متاحين للعب في أيِّ وقت.

وفي دراسةٍ استطلاعية؛ قامت جامعة أوكلاهوما الحكومية بتحليل الإصابات الرياضية بكرة القدم الأمريكية باستخدام التقارير والتحليلات التنبؤية. واتبع المشروع منهجية CRISP-DM (التي سنعرضها في الفصل الرابع)؛ لفهم مشكلة تقديم توصيات بشأن إدارة الإصابات، وفهم عناصر البيانات المختلفة التي تم جمعها حول الإصابات وتنظيف البيانات وتطوير تصويرات لجذب استدلالات مختلفة وبناء تحليلات إدارة الأداء لتحليل فترة التئام الجروح، وقواعد تسلسل الرسم؛ للتنبؤ بالعلاقات بين الإصابات ومختلف أجزاء الجسم المصابة بالإصابات.

وتتألــف مجموعــةُ بيانات الإصابات من أكثر من ٥٦٠ ســجل إصابة في كرة القدم، والتــي تمَّ تصنيفها في إطــار المتغيرات الخاصة بالإصابة إلى: الجزء من الجســم/ موقع الجســم/ الشكل الجانبي، والإجراء المتخذ، والشِّدة، نوع الإصابة، بداية الإصابة وتواريخ الشفاء، والمتغيرات الخاصة باللاعب/ الرياضة إلى: بطاقة تعريف اللاعب، ومركز اللاعب في الملعب، والنشاط، والبداية وموقع المباراة. وتمَّ حسابُ وقت الشفاء من الإصابة لكلِّ سجل؛ إذ تمَّ تصنيفها إلى مجموعات مختلفة من الفترات الزمنية: ٠-١ شهر، ١-٢ شهر، ٢-٤ أشهر، ٤-٦ أشهر، و٦ أشهر- ٢٤ شهرًا.

وتمَّ إنشــاءُ تصويرات مختلفة؛ لاســتنتاج معلومات من مجموعة بيانات الإصابات التي تصــوِّر فترة الشــفاء المرتبطة بمراكز اللاعبين وشــدة الإصابات وفترة الشــفاء، والعلاج المقدَّم والفترة الزمنية للشفاء المرتبطة بها، والإصابات الكبيرة التي تصيبُ أجزاء الجسم، وما إلى ذلك.

وتــمَّ بنــاءُ نمــاذج الشــبكات العصبيــة لــكلِّ فئة مــن فئات الشــفاء باســتخدام IBM SPSS Modeler. وكان مــن المتغيرات المتوقعة: الحالة الراهنة للإصابة، الشــدة، جزء الجسم، موقع الجسم، نوع الإصابة، النشاط، موقع الحدث، الإجراء المتخذ، والموقف. وكان نجاحُ تصنيف فئة الشفاء جيداً جداً؛ إذ كانت الدقة ٧٩,٦٪. وبناءً على التحليل تم اقتراح العديدِ من توصيات الأعمال؛ منها استخدامُ المزيد من مُدخلات المختصين من بداية الإصابة بدلاً من الســماح لموظفي غرفــة التدريب بفحص اللاعبين المصابين. وتمَّ تدريبُ اللاعبين في المراكز الدفاعية لتجنب الإصابة، وممارسة التدريب على آليات فحص السلامة.

أسئلة للمناقشة:

١- ما هي أنواع التحليلات المطبَّقة في تحليل الإصابات؟

٢- كيف تساعد التصويراتُ في فهم البيانات وتقديم رؤى إلى البيانات؟

٣- ما هي مشكلة التصنيف؟

٤- ما الذي يمكن اشتقاقه عن طريق إجراء تحليل تتابعي؟

ما الذي يُمكِننا تعلُّمه من هذه الحالة العملية؟

وبالنســبة لأيِّ مــشروع تحليلات؛ من المهمِّ دائماً فهم نطاق الأعــمال والحالة الراهنة لمشــكلة الأعمال من خلال تحليل مُوسَّــع للمورد الوحيد وهو البيانات التاريخية. وغالباً ما توفر عمليات التصوير المرئي أداةً رائعةً لاكتســاب الأفكار الأوَّلية للبيانات، والتي يمكِن تحســينها بشكلٍ أكبر اســتنادًا إلى آراء الخبراء لتحديد الأهمية الجوهرية لعناصر البيانات ذات الصِّلة بالمشكلة. كما تساعد المعلومات المرئية في توليد أفكار لمشكلات غامضة، والتي

يمكن متابعتها في بناء رؤى إدارة الأداء، والتي يمكن أن تساعد المنظمات في صُنع القرار.

Source: Sharda, R., Asamoah, D., & Ponna, N. (2013). "Research and Pedagogy in Business Analytics: Opportunities and Illustrative Examples." Journal of Computing and Information Technology, 21(3), 171182-.

التحليلات التوجيهية:

تُسمَّى الفئة الثالثة من التحليلات بالتحليلات التوجيهية، والتي تهدف إلى التعرُّف على ما يجري بالإضافة إلى التوقعات المحتملة واتخاذ القرارات؛ لتحقيق أفضل أداء ممكن. وقد تمَّت دراسة هذه المجموعة من الأساليب تاريخيًّا تحت مِظلة التقارير التشغيلية أو علوم الإدارة، وتهدف بشكلٍ عام إلى تحسين أداء النظام. والهدف هنا هو تقديم قرار أو توصية بخصوص إجراء مُحدَّد. وقد تكون هذه التوصيات على شكل (نعم/ لا) لاتخاذ قرار مرتبط بشكلٍ مُحدَّد لكلِّ مشكلة، أو مبلغ معين (على سبيل المثال: سعر بندٍ معين أو سعر الشحن)، أو مجموعة كاملة من خطط الإنتاج. وقد تُعرَض القرارات على صانع القرار في تقرير أو يمكن استخدامها مباشرةً في نظام قواعد القرار الآلي (كما هو الحال في أنظمة تسعير شركات الطيران). وبالتالي؛ يمكن أيضاً تسمية هذه الأنواع من التحليلات بتحليلات القرار أو التحليلات المعيارية. وتُقدِّم الحالة العملية ١,٥ مثالاً على مثل هذه التطبيقات التحليلية المسبقة، وسوف نتعرف على بعض جوانب التحليل التوجيهي في الفصل السادس.

حالة عملية ١-٥

شركة ستيل بار (Steel Bar) المتخصصة تستخدم التحليلات لتحديد تواريخ تنفيذ التزاماتها وفقاً للموارد المتاحة

هذه الحالة العملية مرتبطةٌ بنا جميعاً؛ فقد كانت شركة ستيل بار (Steel Bar) لا ترغب في الكشف عن اسمها (أو حتى مجال عملها الدقيق) تواجه مشكلة كبرى تتمثل في اتخاذ القرارات بشأن مخزون المواد الخام الذي تستخدمه لإرضاء العملاء؛ إذ تقوم هذه الشركة بتوفير قضبان حديدية مُخصَّصة لعملائها. وقد تقوم بتقطيع القضبان إلى أشكال أو أحجام مُحدَّدة، وقد تحتوي على مواد فريدة ومتطلبات تشطيب؛ إذ تقوم الشركة بشراء المواد الخام من جميع أنحاء العالم وتخزينها في مستودعها، وعندما يتصل عميلٌ محتمل بالشركة لطلب عرض أسعار القضبان المتخصصة وما يرتبط بذلك من بيانات (التكوين، أصل المعدن، الجودة، الأشكال، الأحجام... إلخ)، وعادةً ما يكون لدى مندوب المبيعات قليلٌ من الوقت لتقديم مثل هذا السعر بما في ذلك التاريخ الذي يمكن فيه تسليم المنتج وبطبيعة الحال، الأسعار، وما إلى ذلك، ومن ثم يجبُ اتخاذ

قرارات بشأن الوفاء بالالتزامات التاريخية، والتي تُحدِّد في الوقت المناسب التواريخ التي يمكن أن يتمَّ تسليمُ المنتجات التي طلبها العملاء. ويُطلَب من مندوب المبيعات اتخاذُ مثـل هذه القرارات من خلال تحليل التقارير حـول المخزون المتاح للمواد الخام. وقد تكون بعض المواد الخام المتوفرة قد التزم المندوب بالفعل بتسليمها؛ تلبيةً لطلب عميل آخر. وبالتالي؛ قد لا يكون المخزون الموجود في المخزن متاحاً بالفعل. ومن ناحيةٍ أخرى؛ قد تكون هناك مواد خام من المتوقع تسـليمها في المستقبل القريب، ويمكن استخدامها أيضـاً لتلبية طلبٍ مُقـدَّم من هذا العميل المحتمل. وأخيراً؛ قد تكون هناك فرصة أيضاً لتحصيل رسـوم على طلبٍ جديدٍ عن طريق إعـادة تخصيص المخزون الذي تمَّ الالتزامُ به مسـبقاً لتلبية هذا الطلـب الجديد مع تأجيل الطلب الذي تـمَّ الالتزامُ به بالفعل. وبالطبع؛ فإنه يجب أن تسـتند هذه القرارات على تحليلِ التكلفة والعائد لتأخير طلب سـابق. وبالتالي؛ يجب أن يتمكن النظام من سـحب البيانات بشكلٍ لحظي من بيانات المخزون، والطلبات المجمَّعة، والمواد الخام الواردة، وقيود الإنتاج، وما إلى ذلك.

ولدعـم هذه القرارات الصادرة بناءً على الوعـد وفقاً للمتاح؛ تمَّ تطويرُ نظام دعم القـرار DSS الفوري؛ من أجل تخصيصٍ أمثل للمخـزون المتاح ولدعم تحليل ماذا - لو الإضافي. ويسـتخدم نظام دعم القرار مجموعةً من نمـاذج البرامج المختلطة والتي يتمُّ حلُّها باستخدام البرامج التجارية. كما أدرجت الشركة نظام دعم القرار في نظام تخطيط موارد المؤسسة الخاصة بها لتسهيل استخدامها لتحليلات الأعمال.

أسئلة للمناقشة:

١- لماذا تُعَدُّ إعادةُ تخصيص المخزون من عميلٍ إلى آخر؛ قضيةً رئيسةً للمناقشة؟

٢- كيف يمكن أن تساعد أنظمة دعم القرار في اتخاذ هذه القرارات؟

Source: Pajouh Foad, M., Xing, D., Hariharan, S., Zhou, Y., Balasundaram, B., Liu, T., & Sharda, R. (2013). "Available-to-Promise in Practice: An Application of Analytics in the Specialty Steel Bar Products Industry." Interfaces, 43(6), 503517-. http://dx.doi.org/10.1287/inte.2013.0693 (accessed July 2016).

التحليلاتُ المطبَّقة على المجالات المختلفة:

لقد أنتجت تطبيقاتُ التحليلات في قطاعات الصناعة المختلفة العديدَ من المجالات ذات الصِّلة أو على الأقل التعبيرات الشـائعة اليوم. ومن المألوف تقريباً أن نربط كلمة تحليلات بأيِّ صناعة أو بيانات وبالإضافة إلى التصنيف العام لتحليلات النص، والتي تهدف إلى الحصول على قيمة من

النص (سيتم دراستها في الفصل الخامس)، أو تحليلات الويب والتي تهدف إلى تحليل بيانات الويب؛ فقد تمَّ تطويرُ العديد من التحليلات المتعلقة بمشكلة أو صناعة معينة. كما تمَّ تطويرُ التحليلات الخاصة بالعديد من المهن/ المجالات الخاصة بالصناعة أو المسائل الخاصة بالمشكلات. ومن أمثلة هذه المجالات تحليلاتُ التسويق، وتحليلات التجزئة، والنظريات الاحتيالية، وتحليلات النقل، والتحليلات الصحية، والتحليلات الرياضية، وتحليلات المواهب، والتحليلات السُّلوكية، وما إلى ذلك. فعلى سبيل المثال: قدَّم القسم ١-١ عبارة التحليلات الرياضية. ويمكن أيضاً تسمية الحالة العملية ١-١ دراسة حالة في تحليلات شركات الطيران، وسيُقدِّم القسم التالي التحليلات الصحية وتحليلات السوق على نطاقٍ واسع. وبمعنى دقيق يتمُّ تصنيفُ أيِّ تحليل منهجي للبيانات في قطاعٍ معين على أنه تحليلات «(املأ الفراغات)». وعلى الرغم من أن هذا قد يؤدي إلى زيادة حجم مفهوم التحليلات؛ فإنَّ الفائدة هي أن المزيدَ من الناس في صناعات مُحدَّدة سيدركون قوة وقدرات التحليلات. كما يدفع إلى التركيز على المتخصصين الذين يطوِّرون ويطبِّقون مفاهيم التحليلات في قطاع رأسي. وعلى الرغم من أنَّ العديد من الأساليب لتطوير تطبيقات التحليلات قد تكون شائعةً؛ فإنَّ هناك مشكلات فريدة في كلِّ شريحة عمودية تؤثر على كيفية جمع البيانات، ومعالجتها، وتحليلها، وتنفيذ التطبيقات. وهكذا؛ فإنَّ التمييز بين التحليلات المستندة إلى التركيز العمودي؛ مفيدٌ للنموِّ الإجمالي للانضباط.

التحليلات أو علم البيانات:

حتى في الوقت الذي يحظى فيه مفهومُ التحليلات بمزيدٍ من الاهتمام في دوائر الصناعة والأوساط الأكاديمية؛ تمَّ إدخالُ مصطلحٍ آخر وأصبح شائعاً وهو مصطلح علم البيانات، والممارسون لهذا العلم يُطلق عليهم «علماء بيانات»، وأحياناً يُنسَب إنشاء مصطلح علم البيانات إلى باتيل (D. J. Patil) من LinkedIn. وكانت هناك بعض المحاولات لوصف الاختلافات بين مُحلِّلي البيانات وعلماء البيانات (على سبيل المثال انظر: emc.com/collateraliaboutinews/emc-data-science-study-wp.pdf). ومن هذه الآراء أنَّ محللَ البيانات هو مجرد مصطلح آخر للمهنيين الذين كانوا يقومون بعمليات ذكاء الأعمال في شكلِ تجميع البيانات وتنظيفها وإعداد التقارير، وأحياناً بعض التصوير، وتشمل مجموعات مهاراتهم، وشملت مجموعات مهاراتهم إكسل Excel، وبعض معارف SQL، وإعداد التقارير. ويمكنك التعرُّف على هذه القدرات على أنها تحليلات وصفية أو تقارير. وعلى النقيض من ذلك؛ فإنَّ عالم البيانات مسؤولٌ عن التحليل التنبؤي والتحليل الإحصائي، والأدوات التحليلية والخوارزميات الأكثر تقدماً. وقد يكون لديه معرفةٌ أعمق بالخوارزميات وقد تحت تصنيفات مختلفة كالتنقيب في البيانات، أو اكتشاف المعرفة، أو تعلُّم الآلة. وقد يحتاج بعض هؤلاء

الباحثـين إلى معرفة برمجة أكثر عمقاً؛ ليكونوا قادرين عـلى كتابة الكود لتنظيف/ تحليل البيانات في اللغـات الحالية المعتمدة على الويب، مثل لغـة Java أو Python، واللغات الإحصائية مثل R. كـما يحتاج العديدُ من محـترفي التحليلات إلى بناء خبرةٍ كبيرة في النمذجـة الإحصائية، والتجريب، والتحليل. ومـرةً أخرى؛ ينبغي أن يدرك قراؤنا أن هذه المصطلحات تندرجُ تحت مظلة التحليلات التنبؤية والتوجيهية. ومع ذلك؛ تشتمل أيضاً التحليلات التوجيهية على خبرات أكثر أهميةً في تقارير التشغيل بما في ذلك الأمثلية والمحاكاة وتحليل القرار وما إلى ذلك. ومن المرجَّح أن يطلق على أولئك الذين يعملون على هذه الحقول علماء البيانات بدلاً من محترفي التحليلات.

إنَّ وجهة نظرنا في التمييز بين التحليلات وعالم البيانات تكمُنُ في درجة المعرفة والمهارات الفنية أكثر من الوظائف. ويبدو أنَّ برامجَ علوم الحاسوب والإحصاء والرياضيات التطبيقية تفضِّل تسمية علـم البيانات، مع الاحتفاظ بتسـمية التحليلات الخاصة بالمهنيين العاملين بشـكلٍ أكبر. وكمثالٍ آخر على ذلك؛ اقترح المتخصصون في الفيزياء التطبيقية اسـتخدام علم الشـبكة كمصطلحٍ لوصف التحليلات التي تتعلق بمجموعات من الناس؛ كالشـبكات الاجتماعية، وشـبكات سلاسل التوريد، وما إلى ذلك. انظـر: /http://barabasi.com/networksciencebook للاطلاع على كتاب دراسي متطوِّر حول هذا الموضوع.

وإذا صرفنا النظرَ عن الاختلاف الواضح في مجموعة المهارات الذي يمتلكها مَنْ يقومون بإجراء تحليلات وصفية وإعـداد التقارير فقط مقابل أولئك الذين ينخرطون في القيام بأنواع التحليلات الثلاثـة؛ فإن التمييز بـين مصطلح التحليلات وعلم البيانات يصبـح غامضاً. ونلاحظ أن خريجي برامج التحليلات لدينا يكونون مسـؤولين عن مهامّ تتوافق بشـكلٍ أكبر مع مهام علماء البيانات، وليس مجرد فقط تقديم تقارير تحليلية. هذا الكتاب يهدف بوضوحٍ إلى إدخال قدرات ووظائف جميـع أنواع التحليلات، والتي تشـمل علـم البيانات وليس مجرد تقاريـر التحليلات، ومن الآن فصاعداً سنستخدمُ التحليلات وعلم البيانات بشكلٍ متبادل.

أسئلة مراجعة على القسم ١-٥:

١- عرِّف التحليلات.

٢- ما هي التحليلات الوصفية؟ وما هي الأدوات المتنوِّعة المُستَخدَمة في التحاليل الوصفية؟

٣- كيف تختلف التحليلات الوصفية عن التقارير التقليدية؟

٤- ما هو مستودع البيانات؟ وكيف يمكِن لتقنية مستودعات البيانات المساعدة في تمكين التحليلات؟

٥- ما هي التحليلات التنبؤية؟ كيف يمكن للمنظمات استخدامُ التحاليل التنبؤية؟

٦- ما هي التحليلات التوجيهية؟ ما هي أنواع المشكلات التي يمكن حلُّها عن طريق التحليلات التوجيهية؟

٧- عرِّف النمذجة من منظور التحليلات.

٨- هل من الجيد اتباع التسلسل الهرمي للتحليلات الوصفية والتنبؤية قبل تطبيق التحليلات التوجيهية؟

٩- كيف يمكن أن تساعد التحليلات في اتخاذ القرار الموضوعي؟

١-٦ أمثلة على التحليلات في مجالات مختارة:

في الفصول المتنوِّعة لهذا الكتاب سنتطرّق للعديد من أمثلة تطبيقات التحليلات، وهذا هو أحد الطرق الرئيسة المُستَخدَمة في هذا الكتاب (التعرض - Exposure). في هذا القسم سنلقي الضوء على مجالين من المجالات المهمة لتطبيق التحليلات والتي حققت نجاحاً ملحوظاً؛ وهما: الرعاية الصحية، وبيع التجزئة.

تطبيقات التحليلات في الرعاية الصحية - أمثلة Humana:

على الرغم من أنَّ تحليلات الرعاية الصحية تُغطِّي مجموعةً واسعةً من التطبيقات من الوقاية إلى التشخيص إلى العمليات الفعَّالة ومَنْع الاحتيال؛ فإننا نركِّز على بعض التطبيقات التي تمَّ تطويرُها في شركة تأمين صحي رئيسة، ووفقاً لما ذكره موقع الشركة على الإنترنت: «تتكامل إستراتيجية الشركة لتقديم الرعاية، وتجربة الأعضاء، والرؤى الإكلينيكية ورؤى المستهلك؛ بهدف تشجيع المشاركة، وتغيير السلوك، والتواصل الإكلينيكي الاستباقي والتعافي». وينطوي تحقيقُ هذه الأهداف الإستراتيجية على استثمارات كبيرة في تقنية المعلومات بشكلٍ عام، والتحليلات بشكلٍ خاص. إنَّ Brian LeClaire هو نائب الرئيس ورئيس قسم المعلومات في Humana، وهي شركة تأمين رئيسة في الولايات المتحدة. وحاصلٌ على الدكتوراه في MIS من جامعة أوكلاهوما. وقد دافع عن التحليلات كمؤثرٍ تنافسي في Humana كما شارك في إنشاء مركزٍ للتميُّز في التحليلات. ووصف المشاريع التالية بأنها أمثلة على مبادرات Humana في نطاق التحليلات التي يقودها Vipin Gopal رئيس التحليلات الإكلينيكية في Humana.

مثال١: منع حوادث سقوط كبار السن - مدخل تحليلي:

تُعَدُّ الإصابة نتيجةً للسقوط خطراً صحيًّا كبيراً لكبار السن الذين يبلغون من العمر ٦٥ عامًا؛ فهم أكثر من ثلث من يصابون في حوادث سقوط كلِّ عام[١]. إن حوادث السقوط هي أيضاً العامل الرئيس لكلٍّ من الإصابات القاتلة وغير المميتة بالنسبة لكبار السن؛ إذ يؤدي ذلك إلى زيادة خطر الإعاقة بنسبة تصل إلى ٥٠٪[٢].

وتُشكِّل تكاليفُ هذه الإصابات عبئاً كبيراً على نظام الرعاية الصحية بالولايات المتحدة الأمريكية؛ إذ بلغت التكاليفُ المباشرة للسقوط بنحو ٣٤ مليار دولار في عام ٢٠١٣م وحده. ومع ارتفاع النسبة المئوية لكبار السن في الولايات المتحدة؛ فمن المتوقع أن تتزايد حوادث السقوط والتكاليف المرتبطة بها. ووفقاً لمراكز التحكم في الأمراض والوقاية منها (CDC)؛ فإنَّ «حوادث السقوط هي مشكلةٌ صحيةٌ عامة يمكن منعها إلى حَدٍّ كبير»[٣].

وتُعَدُّ Humana ثاني أكبر مورد لبرنامج Medicare Advantage الذي يشملُ حوالي ٣,٢ مليون عضو، معظمهم من كبار السن. إنَّ الحِفاظ على أعضائها البارزين ومساعدتهم في العيش بأمان في منازلهم؛ هو هدفٌ تجاري رئيس؛ إذ يُعَدُّ منعُ حوادث السقوط مكسباً مهماً. ومع ذلك؛ لم تكن هناك منهجيةٌ صارمةٌ لتحديد الأفراد الأكثر عرضةً للسقوط والذين تُعَدُّ جهود الوقاية مفيدةً لهم. وعلى عكس الحالات الطبية المزمنة مثل السكري والسرطان؛ فإنَّ السقوط ليس حالةً طبيةً واضحة المعالم. وبالإضافة إلى ذلك؛ لا يتم الإبلاغ عن حالات السقوط عادةً في بيانات المطالبات؛ نظراً لأن الأطباء يميلون في الغالب إلى ترميز نتائج السقوط، مثل: الكسور والاضطرابات. وعلى الرغم من وجود العديدِ من التقييمات التي تتمُّ إدارتها إكلينيكيًّا لتحديدِ وجود المتعثرين؛ فإنها لا تصِل إلا لعددٍ محدودٍ من الحالات وتفتقر إلى طاقة فعلية كافية. وعلى هذا النحو؛ فهناك حاجةٌ إلى طريقةٍ مستقبلية دقيقة لتحديد الأفراد الأكثر عُرضةً لخطر السقوط؛ حتى يتمكنوا من اتخاذ إجراءات استباقية لمنع السقوط. وقد أجرى التحليل الإحصائي لشركة Humana عملية تطوير نموذج تنبؤي بحوادث السقوط. وهذا هو أول تقرير رسمي شامل يستفيدُ من المطالبات الطبية والصيدلية الإدارية، والبيانات الإكلينيكية، والأنماط الإكلينيكية المؤقتة، ومعلومات المستهلك، وغيرها من البيانات لتعريف الأفراد المُعرَّضين لخطر السقوط في نطاق زمني مُحدَّد.

تُعَدُّ طريقة بي أم فالز (PM Falls) حالياً عنصراً أساسياً لقدرة شركة هيومانا (Humana) على تحديد كبار السن الذين بإمكانهم الاستفادة من التدخلات للتخفيف من سقوطهم. تحديد كبار السِّن الذين يمكنهم الاستفادة من تدخلات التخفيف من السقوط. وقد أظهر برهانٌ مبدأيٌّ للمفهوم مع مستهلكي Humana، الذين يمثلون الـ ٢٪ الأعلى ضمن مخاطر السقوط؛ أنَّ عددَ المستفيدين من خدمات العلاج الطبيعي قد ارتفع، مشيراً إلى أنَّ المستهلكين يتخذون خطواتٍ فعالةً لتقليل مخاطر السقوط. وهناك مبادرةٌ ثانيةٌ تستخدم طريقة PM Falls لتحديد الأفراد المُعرَّضين للخطر حتى يخضعوا لبرامج المراقبة عن بُعد. وباستخدام PM، تمكَّنَت Humana من تحديد ٢٠,٠٠٠ مستهلك تعرَّضوا لخطر السقوط، وهم الذين استفادوا من هذا البرنامج. ويرتدي المستهلكون الذين تمَّ تحديدهم جهازاً يكتشف السقوط وينبِّه المستهلك طُوال اليوم وعلى مدار الأسبوع للمساعدة الفورية.

وحصـــل هذا العمل على جائزة Analytics Leadership من قِبل كلية Kelly لإدارة الأعمال بجامعة Indiana في عام ٢٠١٥؛ وذلك لتبنِّيه أساليب التحليل في بيئة العمل.

مثال٢: هدف هيومانا الجريء - تطبيق التحليلات لتحديد المقاييس المناسبة:

في عـام ٢٠١٤م؛ أعلنت شركة Humana عن هدف المنظمة الجريء لتحسـين صحة المجتمعات المحلية بنسبة ٢٠٪ بحلول عام ٢٠٢٠م من خلال تسهيل الأمر لتحقيق أقصى درجات الصحة. ويمكِن تعريف المجتمعات التي تخدمها Humana بطرق عديدة، جغرافياً (الولاية، المدينة، الحي)، وحسب المنتج (برنامج Medicare Advantage، الخطط القائمة على صاحب العمل، الخدمات المشتراة بشكلٍ فردي)، أو عن طريق الملف الإكلينيكي (شروط الأولوية بما في ذلك مرض السكري، ارتفاع ضغط الدم، CHF [فشـل القلب الاحتقـاني]، CAD [مرض الشريان التاجي]، COPD (مرض الانسـداد الرئوي المزمن) أو "الاكتئاب". إنَّ فهم صحة هذه المجتمعات وكيفية تتبُّعها بمرور الوقت؛ أمرٌ بالغ الأهمية ليس فقط لتقييم الهدف؛ ولكن أيضاً في صياغة إستراتيجيات لتحسين الصحة العضوية في مجملها.

وكان التحَدِّي أمام منظمة التحليلات يتمثل في تحديد مقياسٍ يُبرز جوهر الهدف الجريء؛ لأن مقاييس التأمين الصحي التقليدية المقاسة موضوعيًّا، مثل القبول في المستشفيات أو زيارات غرف الطـوارئ لكلِّ ١٠٠٠ شـخص لن تنهض هذه المهمة الجديـدة. وكان الهدف هو تحديدُ المقياس الـذي يحافظ عليها ويُحسِّـنها في المجتمـع؛ ولكن يجب أن يكون مرتبطـاً بأعمال شركة هيومانا (Humana). ومـن خلال تقييمات تحليلية صارمـة؛ اختارت Humana في نهاية المطاف "الأيام الصحية"؛ وهو عبارة عن اسـتبيان من أربعة أسـئلة، عن جودة الحياة، وقد تمَّ تطويره من قِبل مركز التحكم في الأمراض لتتبع وقياس تقدُّمهم نحو تحقيق الهدف الجريء.

كان من الضروري التأكُّد من أن المقياس المُحدَّد يرتبط ارتباطاً كبيراً بمقاييس الصحة والأعمال؛ بحيث يؤدي أيُّ تحسُّـن في الأيام الصحية إلى تحسـين الصحة، وإلى نتائج أعمال أفضل، وتتضمَّن الأمثلة التالية كيفية ارتباط "الأيام الصحية" بمقاييس الاهتمام:

- يظهر الأفراد الذين لديهم أيام غير صحية أكثر اسـتخداماً وأنماط تكلفة عالية. لكلِّ ٥ أيام غير صحية إضافية هناك:

(أ) زيادة قدرها ٨٢ دولاراً في متوسط التكاليف الطبية والصيدلية الشهرية.

(ب) زيادة قدرها ٥٢ دخولاً للمستشفيات لكلِّ ١٠٠٠ مريض.

(جـ) زيادة ٠,٢٨ يوماً في متوسط مدة البقاء.

- الأفراد الذين يُظهرون سـلوكيات سـليمة وظروفهم المزمنة تُدار بشكلٍ جيد لديهم عدد أيام غير صحية أقل، على سـبيل المثال: الأفراد المصابون بمرض السـكري لديهم عددُ أيام غير صحية أقل إذا حصلوا على فحص الدهون السيئة أو فحص العين السُكرية وبالمثل، إذا كان هناك تحكم في مستويات السكر في الدم.
- الأفـراد الذين يعانون مـن الأمراض المزمنة لديهم عدد أيام غير صحيـة أكثر ممَّن ليس لديهم حالات مزمنة ومنها: مرض السكري، ارتفاع ضغط الدم، CHF [فشل القلب الاحتقاني]، CAD [مرض الشريان التاجي]، COPD (مرض الانسداد الرئوي المزمن) "أو الاكتئاب".

ومنذ ذلك الحين، اعتمدت Humana الأيام الصحية كمقياس لها لقياس التقدُّم نحو تحقيق هدفها الجريء.

مثال٣: نماذج تنبؤية لتحديد أعلى نسبة من المخاطر في شركة تأمين صحي:

تنطبـق قاعدة ٨٠/٢٠ بشـكلٍ عام عـلى الرعاية الصحية، وتعني أنَّ حـوالي ٢٠٪ من المرضى يستهلكون ٨٠٪ من موارد الرعاية الصحية؛ بسبب حالتهم الصحية المتدهورة أو المزمنة.

إنَّ تحديد الأعضاء المناسـبين أمرٌ حاسمٌ في التسجيل في الرعاية الصحية، وفي السنوات الأخيرة تمَّ تطويرُ أنظمةٍ لإدارة الأداء تقوم بتحديد المُسجَّلين الأعلى مخاطر في المستقبل. كما تمَّ تطويرُ العديد من أنظمة إدارة الأداء هذه مع الاعتماد الكبير على بيانات المطالبات الطبية، والتي تنتج عن الخدمات الطبية التي يسـتخدمها المسجلون. بسـبب التأخر الموجود في تقديم بيانات المطالبات ومعالجتها، هنـاك تأخر مماثل في تحديد الأعضاء المعرضين لمخاطر عالية للتسـجيل في البرامج السريرية، وهذه المسـألة ذات أهميةٍ خاصةً عندما ينضمُّ أعضاء جُدُد إلى شركة تأمين صحي كون ليس لديهم تاريخ مطالبات صحي مع شركات تأمين أخرى. يمكن أن يستغرق مُقدِّم الطلب الذي يعتمدُ على المطالبات في المتوسط ٩-١٢ شهراً بعد تسجيل أعضاء جُدُد لتحديدهم للإحالة إلى البرامج الإكلينيكية.

في أوائـل العِقد الجاري جذبت شركـة هيومانا العديدَ من الأعضاء الجُـدُد من خلال منتجات ميدكير إدفانتج (Medicare Advantage)، وبالتالي احتاجت إلى طرقٍ أفضل لإدارة الأعضاء. وعلى هـذا النحو؛ أصبح من المهمّ للغاية تطوير نهج تحليلي آخـر لتحديد الأعضاء الجدد ذوي المخاطر العالية بشكلٍ سريعٍ ودقيق للإدارة الإكلينيكية، للحفاظ على صحة هذه المجموعة وخفض التكاليف.

وقـد قام فريق التحليـلات الإكلينيكية التابع لـ Humana بتطوير نمـوذج التنبؤ بالأعضاء الجدد (NMPM) والذي سيُحدِّد بسرعة الأفراد المعرَّضين للخطر، وسيغيِّر قريباً تسجيلهم الجديد للبرنامج مع Humana، بدلاً من الانتظار حتى يصبح تاريخ المطالبة الكافي متاحاً لتجميع الملفات الإكلينيكيـة والتنبؤ بمخاطر الصحة المسـتقبلية. وقد تمَّ تصميمُ نموذج التنبـؤ بالأعضاء الجُدُد

لمواجهة التحَدِّيات الفريدة المرتبطة بالأعضاء الجُدُد، وهو أسلوبٌ جديدٌ استفاد من مجموعات البيانات الأوسع نطاقاً ودمجها بخلاف بيانات المطالبات الطبية، مثل بيانات تقييم المخاطر الصحية التي تم الإبلاغ عنها ذاتياً والمؤشرات المبكرة من بيانات الصيدلة، واستخدام أساليب التنقيب في البيانات المتقدِّمة لاكتشاف النمط، وسِجل المستهلك يوميًّا في Medicare Advantage استنادًا إلى أحدث بيانات Humana حتى الآن. وقد تمَّ نشرُ النموذج من خلال فريقٍ مُتعدِّد الوظائف من التحليلات وتقنية المعلومات والعمليات؛ لضمان تكامل التشغيل وتكامل الأعمال بسلاسة.

ومنذ أن تمَّ تطبيقُ نموذج التنبؤ بالأعضاء الجدد في يناير ٢٠١٣؛ فقد استطاعوا بشكلٍ سريعٍ تحديدَ الأعضاء ذوي المخاطر العالية من أجل تسجيلهم في برامج تعليمية إكلينيكية. وقد تمَّ إبرازُ النتائج الإيجابية التي تحققت من خلال هذا النموذج في العديد من اتصالات القيادات العليا من Humana. وفي الربع الأول من عام ٢٠١٣، أعلن **Bruce Broussard**، الرئيس التنفيذي لشركة Humana، عن "الزيادة التي طرأت على الأعضاء الجدد والتحسُّن في عمليات التقييم الإكلينيكي"، والتي أدَّت إلى تسجيل ٣١٠٠٠ عضو جديد في البرامج الإكلينيكية، مقارنةً بـ ٤٠٠٠ عضو في نفس البرنامج عن فترة سابقة من العام، بزيادة قدرها ٦٧٥٪. بالإضافة إلى زيادة حجم التسجيل في البرنامج الإكلينيكي؛ أظهرت دراسات النتائج أن المستهلكين الجُدُد المسجَّلين الذين حَدَّدهم نموذج التنبؤ بالأعضاء الجدد قد تمَّ تحويلهم إلى البرامج الإكلينيكية في وقتٍ أقرب، مع أكثر من ٥٠٪ من الحالات المحوَّلة التي تم تحديدها خلال الأشهر الثلاثة الأولى بعد التسجيل الجديد في Medicare Advantage، وقد شارك المستهلكون الذين تمَّ تحديدهم أيضاً بمعدل أعلى واستمروا لفترة أطول في البرامج.

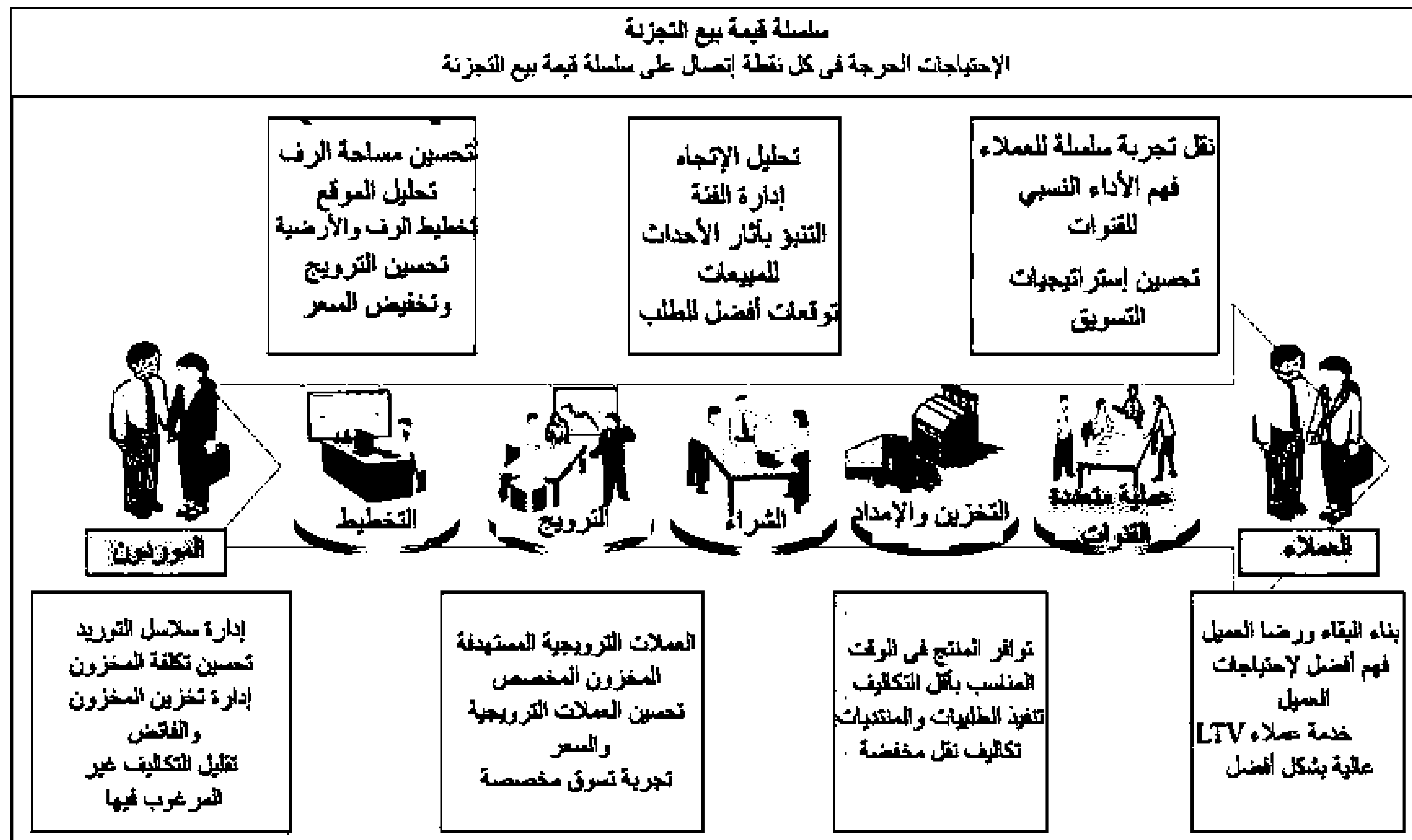
توضِّح هذه الأمثلة كيفية قيام المنظمة باستكشاف تطبيقات التحليل وتطبيقها لتحقيق أهدافها الإستراتيجية. وسنعرض في الفصول القادمة من هذا الكتاب العديد من الأمثلة الأخرى لتطبيقات الرعاية الصحية.

تحليلات سلسلة قيمة البيع بالتجزئة:

يُعَدُّ قطاع البيع بالتجزئة؛ هو المكان الذي قد ترى فيه معظم تطبيقات التحليلات، وهو مجالٌ تكون فيه القيم كبيرةً ولكن هوامش الربح عادةً تكون ضئيلةً، وفيه تتغير أذواق العملاء وتفضيلاتهم بشكلٍ متكرِّر؛ مما يجعل المتاجر - المباشرة أو عبر الإنترنت - تواجه العديدَ من التحَدِّيات في سبيل النجاح. وهيمنة السوق في وقتٍ واحدٍ لا تضمن نجاحًا مستمرًا؛ لذا فإنَّ الاستثمار في تعلُّم الكثير عن الموردِّين والعملاء والموظفين وكلِّ المساهمين الذين يدفعون سلسلة قيمة بيع التجزئة إلى النجاح واستخدام تلك المعلومات لاتخاذ قراراتٍ أفضل كان هدفاً لصناعة التحليلات لفترة طويلة. وحتى قراء التحليلات العارضين يعرفون استثمارات Amazon الهائلة في التحليلات لتقوية سلسلة القيمة التابعة لهم. وبالمثل؛ استثمرت كلٌّ من Walmart وTarget وغيرهما من كبار تجار التجزئة

ملايــين الــدولارات في التحليلات الخاصة بسلاســل التوريد الخاصة بهــم. ويتمتع معظم مورِّدي تقنيــة وخدمة التحليلات بحضورٍ كبيرٍ في تحليــلات بيع التجزئة، وتغطية ولو جزء صغير من هذه التطبيقات لتحقيق هدف التعرُّض الذي نحن بصدد دراسته يمكن أن تملأ كتابًا كاملًا. لذلك يُسلِّط هذا القســم الضوء فقط على بعــض التطبيقات المحتملة. وقد تمَّ إجــراءُ معظم هذه العمليات بواســطة العديد من تجــار التجزئة ويتمُّ توفيرها مــن خلال العديد من مُــورِّدي التقنية؛ ولذلك ففي هذا القســم ســوف نلقي نظرةً أكثر عموميةً بدلًا من الإشارة إلى حالات مُحدَّدة. وتمَّ اقتراحُ هــذه النظرة العامة من قِبل Abhishek Rathi، المدير التنفيذي لـ vCreaTek.com، وهي شركةٌ لخدمات البرمجيات التحليلية ولديها مكاتبُ في الهند والولايات المتحدة والإمارات العربية المتحدة وبلجيكا؛ إذ تقوم الشركة بتطوير تطبيقات في مجالات متعددة، وخصوصًا تحليلات بيع التجزئة.

ويعرض الشكل (١-١٢) المكونات المختارة لسلسلة قيمة بيع التجزئة. يبدأ الشكل بالمورِّدين ويختتم بالعملاء؛ ولكنه يوضِّح العديدَ من نقاط قرار التخطيط الإستراتيجي والتشغيلي المتوسط؛ إذ يمكــن أن تلعبَ التحليلات - الوصفيــة أو التنبؤية أو التوجيهية - دورًا في اتخاذ قراراتٍ أفضل تعتمد على البيانات. كما يوضح الجدول (١-١) أيضاً بعض المجالات المهمة لتطبيقات التحليلات، وأمثلة على الأســئلة الرئيســة التي يمكن أن تجيب عنها التحليلات، وبــكلِّ تأكيد، قيمة الأعمال المحتملة المشتقة من إجراء مثل هذه التحليلات. ثم يتم مناقشة بعض الأمثلة فيما بعد.



شكل ١-١٢: مثال لتطبيقات التحليلات في سلسلة قيمة بيع التجزئة

جدول ١-١: أمثلة لتطبيقات التحليلات في سلسلة قيمة بيع التجزئة

تطبيق التحليل	سؤال الأعمال	قيمة الأعمال
تحسين المخزون	١- أيُّ المنتجات عليها طلبٌ أعلى؟ ٢- ما هي المنتجات بطيئة الحركة أو المتقادمة؟	١- توقع استهلاك المنتجات سريعة الحركة، وطلب مخزونٍ كافٍ منها؛ لتجنُّب سيناريو نفاذ المخزون. ٢- رفع مُعدَّل دوران المخزون للمنتجات بطيئة الحركة؛ من خلال الجمع بينها في طلب مرتفع.
مرونة السعر	١- ما مقدار صافي هامش الربح؟ ٢- ما مقدار الخصم الممكن منحه على المنتج؟	١- تخفيض السعر لكلِّ منتج يمكن أن يقلل خسارة هامش الدولار. ٢- يتمُّ تحديدُ السعر الأمثل لحزمة المنتجات لتوفير هامش الدولار
تحليل سلة السوق	١- ما هي المنتجات التي يجب دمجها لإنشاء عرض حزمة؟ ٢- هل يجبُ الجمع بين المنتجات بناءً على خصائص الحركة البطيئة والسريعة الحركة؟ ٣- هل يجب إنشاء حزمة من نفس الفئة أو خط فئة مختلف؟	١- يُحدِّد تحليلُ التقارب الارتباطات الخفية بين المنتجات، والتي يمكن أن تساعد في القيم التالية: • وَضْع إستراتيجية لحزمة المنتج؛ استنادًا إلى التركيز على المخزون أو الهامش. • زيادة عمليات cross-sell عن طريق إنشاء حزمة من فئات مختلفة أو زيادة عمليات up-sell من نفس الفئات.
رؤية المتسوق	مَنْ من العملاء يشتري أيَّ المنتجات في أيِّ مكان؟	عن طريق تقسيم العملاء؛ يمكن لمالك الأعمال إنشاء عروض مُخصَّصة تحسِّن تجربة العملاء، وتؤدي إلى الاحتفاظ بهم.

تطبيق التحليل	سؤال الأعمال	قيمة الأعمال
تحليل تحوُّل العميل إلى المنافسين	١- مَنْ هم العملاء الذين لن يعودوا؟ ٢- ما مقدار الخسارة في الأعمال؟ ٣- كيف يمكِن الاحتفاظ بهم؟ ٤- ما التركيبة السكانية للعميل صاحب الولاء؟	١- يمكِن تحديدُ العلاقات بين العملاء والمنتجات التي تؤدي لتحوُّل العملاء. وبالتالي يمكِن أن يتمَّ التركيزُ بشكلٍ أفضل على جودة المنتج والسبب وراء هذا التحول. ٢- بالاعتماد على قيمةِ حياة العميل (LTV)؛ يمكن القيامُ بالتسويق الموجَّه؛ مما يؤدي إلى الاحتفاظ بالعميل.
تحليل القنوات	١- ما هي قناة اكتساب العملاء ذات التكلفة الأقل؟ ٢- ما هي قناة الاحتفاظ الأفضل بالعملاء؟ ٣- ما هي القناة الأكثر ربحية؟	يمكِن تحسينُ ميزانية التسويق بناءً على رؤية العائد الأفضل على الاستثمار.
تحليل الفروع الجديدة	١- ما هو الموقع الواجب اختياره؟ ٢- ما هو المخزون الافتتاحي الواجب اقتناؤه وما مقداره؟	١- يمكِن استخدامُ أفضل الممارسات للمواقع والقنوات الأخرى للبدء في القفز. ٢- يمكِن أن تساعدَ المقارنة ببيانات المنافس في إنشاء عامل تمييز/ USP لجذب العملاء الجدد.
تصميم المتجر	١- كيف يجبُ تخطيط المتجر؛ من أجل الخطوط العليا الأفضل؟ ٢- كيف يمكنُ زيادة تجربة العملاء داخل المتجر؟	١- فهم ارتباط المنتجات لاتخاذ قرارٍ بشأن تصميم المتجر وتوافقه بشكلٍ أفضل مع احتياجات العملاء. ٢- يمكِن تخطيطُ نشر القوى العاملة لتفاعل العملاء بشكلٍ أفضل وبالتالي إرضاء تجربة العملاء.

تطبيق التحليل	سؤال الأعمال	قيمة الأعمال
تحليلات الفيديو	١- ما التركيبة السُّكانية التي تدخل المتجر خلال فترة ذروة المبيعات؟ ٢- كيف يمكن تحديدُ العميل صاحب أعلى (LTV) عند مدخل المتجر؛ بحيث يمكن توفير تجربة شخصية أفضل لهذا العميل؟	١- يمكن تخطيطُ العروض الترويجية والأحداث داخل المتجر بناءً على التركيبة السكانية للزيارات الواردة. ٢- مشاركة العميل المُستهدَف والخصم الفوري يُعزِّزان تجربة العميل؛ مما يؤدي إلى زيادة الاحتفاظ به.

عادةً ما يَعرف موقع بيع التجزئة عبر الإنترنت عملاءه بمجرد تسجيل دخول العميل، وبالتالي يمكنهم تقديم صفحات/ عروض مُخصَّصة لتحسين التجربة. وبالنسبة لأيِّ متجر لبيع التجزئة؛ فإنَّ معرفة عميله عند مدخل المتجر لا يزال يُمثل تحديًا كبيرًا. ومن خلال الجمع بين تحليلات الفيديو والمعلومات/ الشَّارات الصادرة من خلال برنامج الولاء، قد يتمكَّن المتجرُ من تحديد العميل عند المدخل نفسه، وبالتالي تمكين فرصة إضافية لعمليات cross-sell أو up-sell. وعلاوةً على ذلك؛ يمكن توفير تجربة تسوق شخصية بمشاركة أكثر تخصيصاً خلال وقت العميل في المتجر.

ويستثمر تجارُ التجزئة الكثيرَ من المالِ في نوافذ عرض جذَّابة أو فعاليات ترويجية أو رسومات مُخصَّصة أو تزيين المتجر أو إعلانات مطبوعة أو لافتات. ولتعزيز فعالية هذه الأساليب التسويقية؛ يمكِن للفريق استخدامُ تحليلات المتسوقين من خلال مراقبة صور الدوائر التلفزيونية المغلقة (CCTV) لمعرفة التفاصيل الديموغرافية لحركة القدم داخل المتجر، ويمكِن تحليلُ صور CCTV باستخدام خوارزميات مُتقدِّمة لاستخراج تفاصيل ديموغرافية، مثل: السِّن والجنس، ومزاج الشخص الذي يتجول بالمتجر.

إضافةً إلى ذلك؛ يُمكِن لبيانات حركة العميل داخل المتجر عند دمجها مع تصميم الرفوف أن تعطي نظرةً ثاقبةً لمدير المتجر لتحديد مناطق البيع السَّاخنة/ المربحة داخل المتجر. وعلاوةً على ذلك؛ يمكِن لمدير المتجر استخدامُ هذه المعلومات لتخطيط توزيع القوة العاملة لهذه المناطق في أثناء فترات الذروة.

وعادةً ما يتمُّ استخدامُ تحليل سلة السوق من قِبل مديري الأقسام لدفع بيع SKUs بطيئة الحركة. وباستخدام التحليلات المتقدِّمة من البيانات المتاحة؛ يمكن أن يكون تقاربُ المنتج في أدنى مستوى من SKU لدفع عائدات الاستثمار على عروض الحزمة لأعلى. وبالإضافة إلى ذلك؛

فباستخدام أســاليب مرونة السعر؛ يمكِن أيضاً تخفيضُ السعر لعرض الباقة؛ مما يقلل أيَّ خسارةٍ في هامش الربح.

وهكــذا، وباســتخدام بيانات التحليلات، لا يمكــن لتجار التجزئة الحصول عــلى معلومات عن عملياتهــم الحالية فحســب؛ بل يمكِن أيضــاً الحصولُ على مزيدٍ من الإحصــاءات لزيادة الإيرادات وخفض التكلفة التشــغيلية لربح أعلى. ويقترح مــدوِّن في علوم البيانات المركزية (Data Science Central) قائمةً شاملةً إلى حَدٍّ ما لتطبيقات تحليلات التجزئة الحالية والمحتملة، والتي يمكِن لمتاجر التجزئة الكبرى مثل Amazon استخدامها. وهذه القائمة متاحةٌ على:

http://www.datasciencecentral.com/profiles/blogs/20-data-science-systems-used-by-ainazon-to-operate-its-business.

وكما ذكرنا ســابقاً؛ فإنَّ هناك العديدَ من الأمثلة على هذه الفرص المُدرَجة هنا؛ ولكنك سترى العديدَ من الأمثلة على مثل هذه التطبيقات في جميع أنحاء الكتاب.

أسئلة مراجعة على القسم ١-٦:

١- لماذا تستثمرُ شركة التأمين الصحي في التحليلات عقب اكتشاف الاحتيال؟ ولماذا من مصلحتهم التنبؤ باحتمالية سقوط المرضى؟

٢- ما هي التطبيقات الأخرى المشابهة للتنبؤ بالسقوط التي يمكنك تصورها؟

٣- كيف تقنع عميل تأمين صحي جديداً بتبَنِّي أنماط حياة صحية (مثال ٣ Humana)؟

٤- حَــدِّد على الأقل ثلاث فرصٍ أخرى لتطبيق التحليلات في سلســلة قيمة البيع بالتجزئة بخلاف تلك التي تمَّ تغطيتها في هذا القسم.

٥- مــا هــي متاجرُ بيــع التجزئة التــي تعرفها التي تســتخدمُ بعض تطبيقــات التحليلات، التي تمَّ تحديدُها في هذا القسم؟

١-٧ مقدمة موجزة لتحليلات البيانات الضخمة:

ما هي البياناتُ الضخمة؟

أيُّ كتــابٍ عن التحليلات وعلم البيانات؛ يجبُ أن يشــمل تغطيةً كبيرةً لما يُســمَّى بتحليلات البيانات الضخمة. وســنقوم بتغطيتها في الفصل السابع؛ ولكننا هنا نعرضُ مقدمةً موجزةً للغاية

عن بداية ظهورها. إنَّ عقولنا تعملُ بسرعة وفعالية عالية ومتنوعة في معالجة كمياتٍ كبيرة من جميـع أنواع البيانات: الصـور، والنصوص، والأصوات، والروائح، ومقاطـع الفيديو. حيث تقوم بمعالجة جميع أشـكال البيانات المختلفة بسـهولةٍ نسـبية. غير أنَّ أجهزة الحاسوب لا تزال تجدُ صعوبةً في مواكبة وتيرة إنشـاء البيانات، فضلاً عن تحليلها بسرعة، وهذا هو السـبب في أن لدينا مشكلة البيانات الضخمة. إذاً، ما البياناتُ الضخمة؟ ببساطة البيانات الضخمة؛ هي البيانات التي لا يمكن تخزينها في وحدة تخزين واحدة. وتشير البياناتُ الضخمة عادةً إلى البيانات التي تأتي في العديدِ من الأشـكال المختلفة: مهيكلةً وغير مهيكلة، في تيار، وما إلى ذلك. إن المصادر الرئيسـة لمثل هذه البيانات هي مقتطفات من مواقع الويب، والمنشورات بمواقع التواصل الاجتماعي مثل Facebook، وبيانات من الزيارات أو من أجهزة الاستشعار أو الطقس. ويحتاج أيُّ محرك بحثٍ على الويب، مثل Google، إلى بحث وفهرسـة المليـارات من صفحات الويب لمنحك نتائج بحث ملائمة في جزءٍ من الثانية. وعلى الرغم من أن هذا لا يتم بشكلٍ لحظي؛ فإنَّ إنشاء فهرس لجميع صفحات الويب على الإنترنت ليس مهمة سهلة. ولحُسْن الحظ؛ فقد تمكَّنت شركة Google، من حل هذه المشكلة. فمن بين الأدوات الأخرى، استخدمت أساليب تحليل البيانات الضخمة.

وهنـاك جانبان لإدارة البيانات بهذا الحجم، وهما: التخزين والمعالجة. فإذا اسـتطعنا شراء حـلِّ تخزين مُكلِّـف للغاية لتخزين كلِّ هذا في مكانٍ واحد على وحـدةٍ واحدة؛ فإنَّ جعل هذه الوحدة تسـمح بالخطأ قد يكون مكلفاً للغاية. ومن ثم فقد تمَّ اقتراحُ حلٍّ عبقريٍّ يشـتملُ على تخزيـن هذه البيانات في أجزاء مختلفة من أجهزة متصلة بشـبكة، ووَضْع نسـخة أو اثنتين من هذه البيانات في مواقع مختلفة على الشـبكة، من الناحيتين المنطقية والمادية. وتمَّ استخدامه في الأصل في Google (والذي كان يُسمَّى آنذاك بنظام ملفات Google)، وتمَّ تطويره وإصداره لاحقاً كمشروع Apache باسم نظام الملفات المقسمة من (Hadoop HDFS).

ومع ذلك؛ فإن تخزين هذه البيانات يمثل نصف المشكلة فقط. فالبيانات لا قيمة لها إذا كانت لا تقـدِّم قيمةً للأعمال، ولكي توفـر قيمةً للأعمال؛ فلا بد من تحليلهـا. ولكن كيف يمكن تحليلُ هـذا القدر الهائل من البيانات؟ حيث لا يمكِن تمريرُ كافة الحسـابات إلى جهاز حاسـوب واحد. فهذا من شـأنه إنشـاءُ قدرٍ من أحمال البيانات على هذا الجهاز. وقد تمَّ اقتراحُ حلٍّ عبقريٍّ آخر. وهو دفع الحوسـبة إلى البيانات، بدلاً من دفع البيانات إلى عقدة الحوسبة. وقد كان هذا نموذجًا جديدًا وأسفر عن طريقة جديدة تمامًا لمعالجة البيانات، وهذا ما نعرفه اليوم باسم نموذج برمجة MapReduce، الـذي جعل معالجة البيانات الضخمة حقيقـة. وقد تمَّ تطويرُ MapReduce في الأصل في Google. وظهر إصدارٌ لاحق من قِبل مشروع Apache يُسمَّى Hadoop MapReduce.

واليــومَ عندما نتحدثُ عن التخزين أو المعالجة أو تحليل البيانات الضخمة؛ فإننا نذكر HDFS وMapReduce عند مستوى معين، وقد تمَّ اقتراحُ معايير وبرامج أخرى ذات صِلة. وعلى الرغم من أن مجموعة الأدوات الرئيســة متاحة بالمجان؛ فقد تمَّ إطلاقُ العديد من الشركات لتقديم التدريب أو الخدمات المتخصصة في هذا المجال. مثل HortonWorks وCloudera وTeradata Aster.

وعلى مدى السنوات القليلة الماضية حدث مزيدٌ من التغير؛ فما كان يُسمَّى البيانات الضخمة أصبــح الآن تطبيقات البيانات الضخمة. وقد أضافت الحاجةُ إلى معالجة البيانات القادمة سرعةً إلى المعادلــة. ومن الأمثلة على معالجة البيانات السريعة؛ التداول الخوارزمي. والذي يســتخدم المنصــات الإلكترونية القائمة على خوارزميات تداول الأســهم في الســوق الماليــة، والتي تعمل في غضون الميكروثانية. ومن الأمثلة الأخرى على المجموعة الواسعة من البيانات؛ هو تحليل المشاعر، والذي يســتخدم أشكالاً مختلفةً من البيانات من مِنصات وسائل التواصل الاجتماعي واستجابات العمــلاء لقياس المشــاعر. واليومَ، ترتبط البيانــات الضخمة بأيِّ نوعٍ من البيانــات الكبيرة التي تتميــزُ بخصائص الحجم والسرعة والتنوع. وتوضِّح الحالة العملية ١-٦ تطبيق تحليلات البيانات الضخمة في صناعة الطاقة. وسوف ندرس تقنيات وتطبيقات البيانات الضخمة في الفصل السابع.

أسئلة مراجعة على القسم ١-٧:

١- ما هي تحليلاتُ البيانات الضخمة؟

٢- ما هي مصادرُ البيانات الضخمة؟

٣- ما هي خصائصُ البيانات الضخمة؟

٤- ما هي المعالجةُ الفنية التي يتمُّ تطبيقها لمعالجة البيانات الضخمة؟

(1) http://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html.

(2) Gill, T. M., Murphy, T. E., Gahbauer, E. A., et al. (2013). Association of injurious falls with disability outcomes and nursing home admissions in community living older persons. American Journal of Epidemiology, 178(3), 418-425.

(3) Gates, S., Smith, L. A., Fisher, J. D., et al. (2008). Systematic review of accuracy of screening instruments for predicting fall risk among independently living older adults. Journal of Rehabilitation Research and Development, 45(8), 1105-1116.

Contributors: Harpreet Singh, PhD; Vipin Gopal, PhD; Philip Painter, MD.

حالة عملية ١-٦

شركة Center Point للطاقة تستخدم التحليلات الفورية للبيانات الضخمة؛ لتحسين خدمة العملاء

تُعَدُّ شركة Center Point للطاقة إحدى الشركات الكبرى لتوصيل الطاقة ويقع مقرُّها في هيوستون، تكساس. وتشمل أعمالها الأساسية نقل وتوزيع الكهرباء، وتوزيع الغاز الطبيعي، ومبيعات وخدمات الغاز الطبيعي. ولديها أكثرُ من خمسة ملايين عميل في الولايات المتحدة.

وتستخدم الشركةُ شبكات الطاقة الذكية؛ لجمع المعلومات الفورية عن صحة الجوانب المختلفة للشبكة، مثل: العدادات والمحولات والمفاتيح المُستَخدَمة في توفير الكهرباء. ويتمُّ تحليلُ هذه المعلومات بشكلٍ فوري باستخدام الطاقة مع تحليلات البيانات الضخمة؛ مما يتيحُ تشخيصًا دقيقًا وأسرع بكثير. فعلى سبيل المثال: يمكن أن تتنبأ وتساعد في منع انقطاع التيار الكهربائي.

إضافةً إلى ذلك؛ تقوم الشركة بجمع بيانات الطقس؛ مما يسمح بأن تساعد البيانات التاريخية في التنبؤ بانقطاع التيار نتيجةً للعاصفة. وتعمل هذه الرؤية كدليلٍ إرشادي لوضع الموارد الصحيحة قبل العاصفة؛ لتجنب حدوث الانقطاع.

ثانياً، ومن أجل أن تفهم عملاءها على نحوٍ أفضل، تستخدم Center Point تحليل المشاعر الذي يفحص رأي العميل عن طريق العاطفة (السعادة والغضب والحزن، وما إلى ذلك)، وتقوم الشركة بتصنيف عملائها بناءً على مشاعرهم ومن ثمَّ تكون قادرةً على البيع لعملائها بطريقة تستند إلى شخصياتهم وتقدم تجارب أكثر قيمة للعملاء.

ونتيجةً لاستخدام تحليلات البيانات الضخمة، وفَّرت الشركة ٦٠٠,٠٠٠ جالون من الوقود في العامين الماضيين؛ من خلال حلِّ ستة ملايين طلب خدمة عن بُعْد. بالإضافة إلى ذلك، وفَّرت الشركة ٢٤ مليون دولار لعملائها في هذه العملية.

أسئلة للمناقشة:

١- كيف يُمكِن لشركات الكهرباء التنبؤ بانقطاع ممكن في موقع؟

٢- ما هو تحليلُ مشاعر العميل؟

٣- كيف يساعدُ تحليلُ مشاعر العملاء الشركات على تقديم خدمة مُخصَّصة لعملائها؟

ما الذي يمكننا تعلُّمه من هذه الحالة العملية:

باستخدام تحليلات البيانات الضخمة؛ يمكن لشركات الطاقة حَلُّ مشكلات العملاء بشكلٍ أفضل مثل انقطاع التيار والأعطال الكهربائية في غضون فترة زمنية أقصر مقارنةً بالعمليات السابقة. كما يمكِن أن يساعد تحليلُ المشاعر في استهداف عملائهم وفقاً لاحتياجاتهم.

Sources: Sap.com, "A 'Smart' Approach to Big Data in the Energy Industry," http://www.sap.com/bin/sapcom/cs_cz/downloadasset.201310--oct-0920-.a-smart-approach-to-big-data-in-the-energyindustry-pdf.html (accessed June 2016); centerpointenergy.com, "Electric Transmission & Distribution (T&D)," http://www.centerpointenergy.com/en-us/Corp/Pages/Company-overview.aspx (accessed June 2016); YouTube.com, "CenterPoint Energy Talks Real Time big Data Analytics," https://www.youtube.com/watch?v=s7CzeSlIEfI (accessed June 2016).

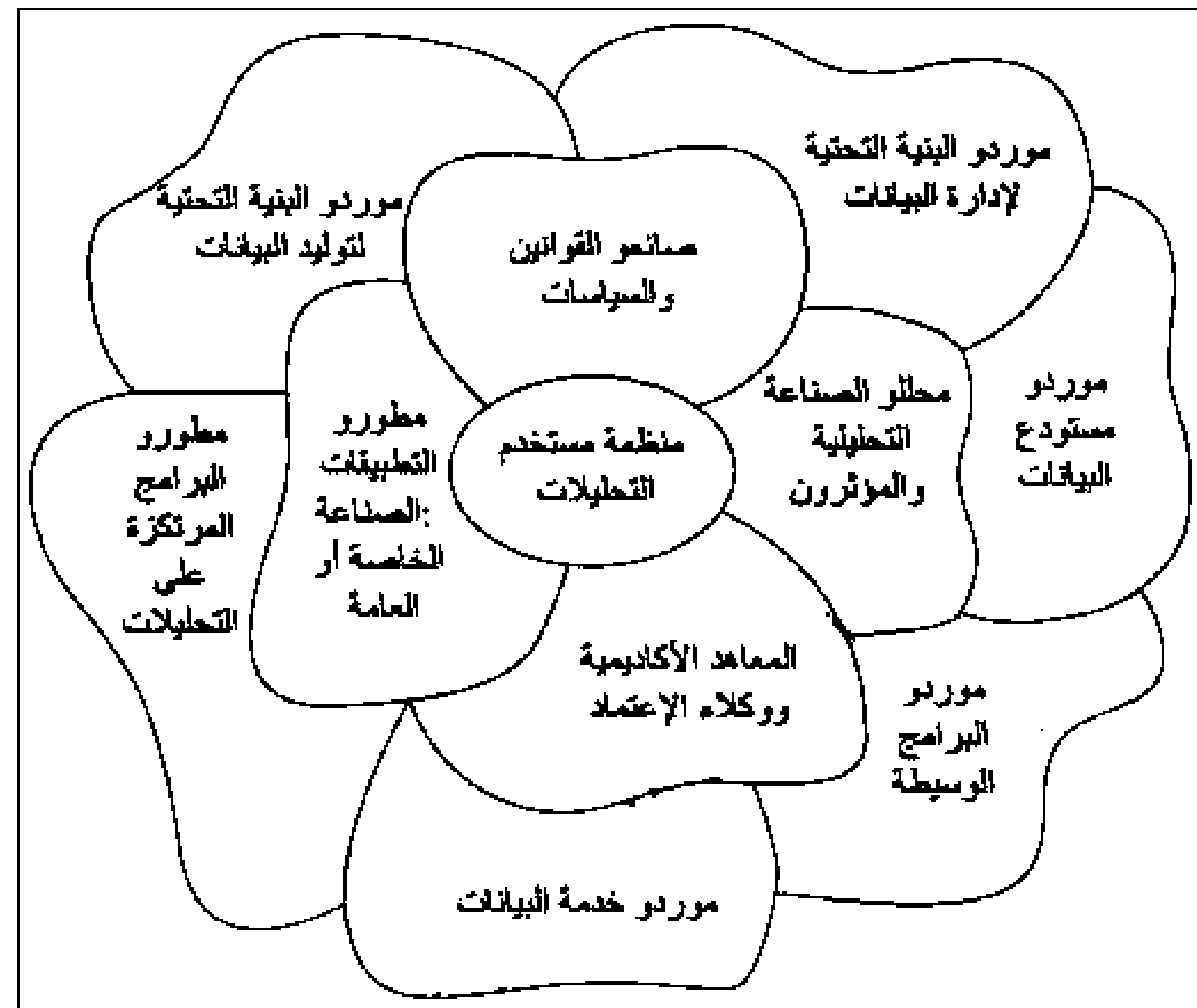
١-٨ نظرة عامة على النظام البيئي للتحليلات:

أنـــت الآن عزيزي القارئ وبناءً على ما تقدَّم؛ مُتحمِّــسٌ لإمكانات التحليلات وتريد الانضمام إلى هــذه الصناعــة المتنامية. ولكن مَنْ هم اللاعبون الحاليون ومــاذا يفعلون؟ وما المكان الذي يناسبك؟ إنَّ الهدفَ من هذا القسم هو تحديدُ قطاعاتٍ مختلفة من صناعة التحليلات، وتقديم تصنيــفٍ للأنــواع المختلفة مـــن المشــاركين في الصناعة، وتوضيــح أنواع الفــرص المتاحة لمحترفي التحليــلات. لقــد تَمَّ تحديدُ أحد عشر نوعًا مختلفًا من اللاعبين في النظام البيئي للتحليلات. كما أنَّ فهــمَ النظام البيئي يعطي القارئ رؤيةً أوســع لكيفية تلاقي اللاعبين المختلفين. وهناك غرضٌ ثانــوي لفهم النظام البيئي للتحليلات، الخاص بمتخصِّــصي ذكاء الأعمال أيضاً، وهو أن يكون على علمٍ بالشركات والعروض والفرص الجديدة في القطاعات المتحالفة مع التحليلات. ويُختتم القسم ببعض الملاحظات عن الفرص المتاحة للمحترفين للتحرُّك عبر هذه المجموعات.

وعــلى الرَّغم مـــن أن بعض الباحثين قد ميَّزوا بين مُتخصِّصي تحليـــل الأعمال وعلماء البيانات (Davenport وPatil، 2012)، كــما أشرنا ســابقاً، بهدف فهم النظام البيئي الشــامل للتحليلات؛ فإننا نعاملهم كمهنة واحدة واسعة. ومن الواضح أن احتياجات المهارات قد تختلف بين أخصائي رياضيــات قــوي إلى مبرمج إلى مصمم نمــاذج إلى أخصائي اتصالات، ونعتقد أنَّ هذه المســألة قد تمَّ حلُّها على مســتوى أصغر/ فردي بدلاً من المســتوى الكلي لفهم مجمــوع الفرص. كما نتبنَّى تعريفًا واسعًا للتحليلات ليشمل جميع الأنواع الثلاثة وفقاً لـ INFORMS وهي الوصفية/ إعداد التقارير/ التصويرات، والتنبؤية والتوجيهية كما هو مُوضَّح سابقًا.

ويُوضِّح الشـكلُ ١-١٣ وجهة نظرٍ واحدة للنظام البيئـي للتحليل. إذ تتمثلُ مكونات النظام البيئـي في بتـلات زهرة التحليـلات. وقد تمَّ تحديدُ أحد عشر قطاعًا أو قسـمًا رئيسًـا في مجال التحليـلات. ويتمُّ تجميع مكونات النظام البيئي التحليلي في ثلاثِ فئات تمثلها البتلات الداخلية، والبتلات الخارجية، وبذرة (الجزء الأوسط من) الزهرة.

ويمكـن أن يُطلَق على البتلات السـتة الخارجية مـوردو التقنية. وتـأتي إيراداتهم من تقديم التقنيـة، والحلول، والتدريب لمنظمات مسـتخدمي التحليلات؛ حتى يتمكَّنـوا من توظيف هذه التقنيـات بأكثر الطـرق فعاليةً وكفاءةً. ويمكِـن تعريفُ البتلة الداخليـة بصفة عامة كمسرعات التحليلات. وتعمل أجهزةُ التسريع مع مورِّدي التقنية والمستخدمين. وأخيراً؛ يشتمل جوهر النظام البيئـي على منظمات مسـتخدمي التحليلات. وهـذا هو العنصر الأكثر أهميـة؛ إذ يتمُّ دفع كلِّ مجموعة من مجموعات صناعة التحليلات من قِبل منظمات المستخدمين.



شكل ١-١٣: النظام البيئي للتحليلات

إنَّ استعارة اسم «زهرة» مناسبٌ تمامًا للنظام البيئي للتحليلات؛ إذ يتداخل العديدُ من الأشخاص بعضهـم مـع بعض. على غِرار كائن حـي مثل الزهرة، حيث تنمو كل هــذه البتلات وتذوب معاً. ونحن نسـتخدمُ مصطلحات المكونـات والمجموعات والبتلات والقطاعات بشـكلٍ متبادلٍ لوصف اللاعبين المختلفين في مجال التحليلات. وسـنقدِّم فيما يلي كلاً من القطاعات الصناعية كما سـنقدِّم بعض الأمثلة للاعبين في كلِّ قطاع. إن قائمة أسـماء الشركات المُدرَجة في أيِّ بتلة ليسـت شـاملة. فقائمة الممثلين للشركات في كلِّ مجموعة تهدف فقط إلى توضيح العرض الفريد للمجموعة لوصف

المواهب التي يمكن استخدامها أو توظيفها. وكذلك؛ فإن ذكر اسم الشركة أو قدرتها في مجموعةٍ مُحدَّدة لا يعني أنها النشاط / العرض الوحيد لتلك المنظمة. والهدف الرئيس؛ هو التركيزُ على القدرات التحليلية المختلفة داخل كلِّ عنصرٍ من عناصر حيِّز التحليلات. وتعمل العديدُ من الشركات في قطاعات متعددة داخل صناعة التحليلات، وبالتالي توفر فرصًا للحركة داخل المجال أفقيًّا ورأسيًّا.

ولقد قام مات تراك (Matt Turck) - وهو شريك برأس المال مع فرست مارك (FirstMark) - بتطوير نظام بيئي للتحليلات يُركِّز على البيانات الضخمة، ويهدف إلى تتبع اللاعبين الجدد والقدامى في مختلف قطاعات صناعة البيانات الضخمة. وتُوجَد صورة مرئية جميلة من تفسيره للنظام البيئي وقائمة شاملة من الشركات على موقعه: http://mattturck.com/2016/02/01/Big-data-landscape/(accessed August 2016) وسنرى أيضاً نظامًا بيئيًا مشابهًا في سياق إنترنت الأشياء (IoT) في الفصل الأخير.

مُورِّدو البنية التحتية لتوليد البيانات:

قد يكون من الأولى أن نبدأ بالتعرُّف على هذه المجموعات من خلال ملاحظة مجموعة جديدة من الشركات التي تمكِّن من توليد وجمع البيانات التي يمكن استخدامها لتطوير رؤى تحليلية. وعلى الرغم من أنَّ هذه المجموعة يمكِن أن تشمل جميع النقاط التقليدية من أنظمة البيع، وأنظمة إدارة المخزون، ومُقدِّمي التقنية لكلِّ خطوة في سلسلة عمليات/ قيمة الشركة، سننظر أساسًا في اللاعبين الجدد حيث كان التركيزُ الأساسي على تمكين المنظمة من تطوير رؤى جديدة لعملياتها بدلًا من تشغيل عملياتها الأساسية. وبالتالي تشمل هذه المجموعة شركات إنشاء بنية تحتية لجمع البيانات من مصادر مختلفة.

ومن المكونات الناشئة لمثل هذه البنية التحتية هو جهاز الاستشعار. حيث تقوم أجهزة الاستشعار بجمع كمية هائلة من البيانات بمعدل أسرع، وقد تمَّ اعتمادُها من قِبل قطاعات مختلفة، مثل: الرعاية الصحية، والرياضة، والطاقة. على سبيل المثال، البيانات الصحية التي تجمعها تستخدم أجهزة الاستشعار بشكلٍ عام لتتبُّع الحالة الصحية للمستخدمين. وبعض أجهزة الاستشعار الرئيسة التي تقومُ بجمع المعلومات الصحية هي AliveCor، وGoogle، وShimmer، وFitbit. وبالمثل؛ فإنَّ صناعة الرياضة تستخدم أجهزة استشعار لجمع البيانات من اللاعبين والميدان لتطوير الإستراتيجيات وتحسين اللعب الجماعي. ومن أمثلة الشركات التي تنتجُ أجهزة الاستشعار ذات الصِّلة بالرياضة Sports Sensors، وZepp، وShockbox، وغيرها. كما تُستَخدَم أجهزة الاستشعار لإدارة الزيارات. ويساعد ذلك في اتخاذ إجراءات فورية للتحكُّم في الزيارات، ومن مورِّدي هذه الخدمة Advantech B+B SmartWorx، وGarmin، وSensys Network.

وتلعبُ أجهزة الاستشـعار دورًا رئيسًا في إنترنت الأشـياء، وتُعَدُّ جزءًا أساسيًّا من الكائنات الذكية. وهي تجعلُ الاتصال بين آلةٍ وأخرى ممكنًا. والجهات الفاعلة الرئيسة في البنية التحتية لإنترنت الأشياء؛ هي إنتل (Intel)، ومايكروسـوفت (Microsoft)، وجوجل (Google)، وآي بي إن (IBM)، وسيسـكو (Cisco)، وسـمارت بـين (Smartbin)، وسـيكو برودكتـس (SIKO Products)، وأوميجا إنجنيرنج (Omega Engineering)، وأبـل (Apple)، وإس إيـه بي (SAP). وربما تكـون هذه هي المجموعة الأكثر تقنيةً في النظام البيئي. وسوف نسـتعرضُ النظام البيئي لإنترنت الأشـياء IoT في الفصل الثامن. وفي الواقع؛ يُوجَد تقريبًا نظامٌ بيئيٌّ عن كلِّ مجموعة من المجموعات التي نُحدِّدها هنا.

مُورِّدو البنية التحتية لإدارة البيانات:

تضمُّ هذه المجموعة جميعَ المنظمات الرئيسـة التي توفر الأجهزة والبرامج التي تفسِّر الأسـاس الجوهـري لجميع حلـول إدارة البيانات. ومن الأمثلة الواضحة على ذلـك كبار الجهات التي توفر البنية التحتية لحوسبة قواعد البيانات، مثل: IBM، وDell، وHP، وOracle، وما إلى ذلك، ومورِّدو حلول التخزين، مثل EMC (والتي اشترتها Dell مؤخرًا) وNetApp؛ والشركات التي تقدِّم مِنصَّات الأجهزة والبرمجيـات المحلية مثل IBM وOracle وتيراداتا، ومورِّدو حلول البيانات الذين يقدِّمون أنظمةً مستقلةً لإدارة قواعد البيانات للأجهزة والبرامج، مثل عائلة SQL Server من مايكروسوفت (Microsoft)، ويندرجُ تحت هـذه المجموعة أيضاً مورِّدو البرامج المتكاملة المتخصِّصة مثل SAP. كما تشـتملُ هـذه المجموعة على منظمات أخرى، مثـل مورِّدي أجهزة قواعـد البيانات ومقدِّمي الخدمات والدامجين والمطورين وغيرهم، ممَّن يدعمون النظم البيئية لهذه الشركات.

وتظهـر العديدُ من الشركات الأخرى باعتبارها لاعبين أساسـيين في مجالات ذات صِلة، وذلك بفضل البنية التحتية للشبكة التي تمكِّن الحوسبة السحابية. وقد اشتهرت شركات مثل (خدمات الويـب مـن Amazon)، و(IBM (Bluemix، وSalesforce.com بتقديم حلول مسـتودعات البيانات والتحليلات الكاملة عبر السَّحابة، والتي تمَّ الآن اعتمادها من الشركات المذكورة.

وهنـاك مجموعـةٌ حديثةٌ من الشركات في مجال البيانات الضخمـة تُعَدُّ هي كذلك جزءًا من هـذه المجموعة. ولا تقوم شركاتٌ مثـل Cloudera وHortonworks بالضرورة بتوفير الأجهزة الخاصة بهم؛ ولكنهم يقدِّمون خدماتِ البنية التحتية والتدريب؛ لإنشاء منصة البيانات الضخمة. وهذا يشـمل مجموعـات Hadoop، وMapReduce، وNoSQL، وSpark، وKafka، وFlume وغيرها من التقنيات المرتبطة بالتحليلات. وبالتالي؛ يمكن أيضاً تجميعهم تحت إشراف استشاريين أو مدربـين لتمكـين البنية التحتية. إنَّ النظـم البيئية الكاملة للاستشـاريين وتكامل البرمجيات، ومُقدِّمـي التدريب وغيرهم من مُقدِّمي خدمات القيمـة المضافة قد أفادت العديد من اللاعبين

الكبـار في مجموعة البنية التحتيـة لإدارة البيانات. وتعمل بعضُ المجموعات المذكورة أدناه على تحديـد هؤلاء اللاعبين؛ لأن العديد منهـم ينتقلُ إلى التحليلات؛ إذ إنَّ الصناعة تحوِّل تركيزها من معالجة المعاملات بكفاءة إلى اشتقاق القيمة التحليلية من البيانات.

مُورِّدو مستودعات البيانات:

توفِّر الشركات التي تركِّز على مسـتودعات البيانات التقنيـة والخدمات التي تهدفُ إلى دمج البيانـات مـن مصادر متعددة. وبالتالي تمكينُ المنظمات من اسـتخلاص وتقديم قيمة من أصول البيانـات الخاصـة بها. وتوفر العديدَ مـن الشركات في هذا المجال الأجهـزة الخاصة بهم لتوفير مسـتودعات البيانات بكفـاءة واسـترجاعها ومعالجتها. تُعَـدُّ شركاتٌ مثـل IBM وOracle و Teradata؛ من اللاعبين الأساسـيين في هذا المجال. وتشـملُ التطورات الأخيرة في هذا المجال أداء التحليلات على البيانات مباشرةً في الذاكرة. وهناك قطاعُ نموٍّ كبير آخر هو مسـتودع البيانات في السَّحابة، ومن أمثلة هذه الشركات Snowflake وRedshift. وتعمل الشركات في هذه المجموعة بوضوح مع جميع الجهات الأخرى في القطاع؛ من أجل توفير حلول وخدمات مسـتودع البيانات ضمن نظامها البيئي وبالتالي تصبح العمود الفقري لصناعة التحليلات. لقد كانت صناعةً رئيسـةً في حَدِّ ذاتها، وبالتالي، مورِّد ومستهلك للمواهب التحليلية.

مُورِّدو البرمجيات الوسيطة:

بـدأ مسـتودعُ البيانات بالتركيز عـلى جلب جميع مخـازن البيانات إلى مِنصة على مسـتوى المؤسسـة. وأصبح إنشـاءُ مغزى لهذه البيانات صناعةً في حَدِّ ذاتها. ويتمثل الهدفُ العام لقطاع البرمجيات الوسيطة في توفير أدوات سهلة الاستخدام للتقارير أو التحليلات الوصفية، والتي تشكِّل جزءًا أساسيًّا من ذكاء الأعمال أو التحليلات المُستَخدَمة في المنظمات. ومن أمثلة الشركات في هذا المجال Microstrategy، Plum، وغيرها الكثير. وقد تَمَّ الاستحواذُ على عددٍ قليلٍ من كبار اللاعبين الذين كانوا وسـطاء مسـتقلين من قِبل الشركات في أول مجموعتين. فعلى سـبيل المثال، أصبحت Hyperion جزءًا من أوراكل واسـتحوذت SAP على Business Objects، واستحوذت IBM على Cognos، وقـد كان هذا القطاعُ مرادفًا لمورِّدي ذكاء الأعمال الذين يقدمون إلى الصناعة خدماتِ لوحة المعلومات والتقارير والتصوير؛ بناءً على بيانات معالجة المعاملات وقاعدة البيانات ومورِّدي مسـتودعات البيانات. وهكذا انتقلـت العديدُ من الشركات إلى هذا المجال على مرِّ السـنين، بما في ذلـك مُورِّدو برامـج التحليلات العامة مثل SAS أو مُقدِّمو خدمـات مرئية جديدة مثل تابلوه (Tableau)، أو العديـد مـن مُقدِّمي التطبيقات المتخصِّصة. ويورد سـجلُ المنتجات في TDWI.

org 201 مــوردًا فقــط في هذه الفئة؛ وذلــك اعتبارًا من يونيــو ٢٠١٦؛ لذلــك كان القطاع قويًّا. (http://www.tdwidirectory.cona/category/business-intelligence-services). ومــن الواضح أنَّ هذا هو القطاع الذي يحاول الانتقال إلى قطاع علم البيانات في الصناعة.

مُورِّدو خدمة البيانات:

يتمُّ توليدُ الكثير من البيانات التي تســتخدمها المنظمة للتحليلات داخليًّا من خلال عملياتها. ولكن هناك العديدُ من مصادر البيانات الخارجية التي تلعب دورًا رئيسًا في اتخاذ القرار بالمنظمات. وتشمل أمثلة مصادر البيانات هذه بيانات التركيبة السكانية، وبيانات الطقس، والبيانات التي تمَّ جمعهــا من قِبل أطراف ثالثة، والتي يمكن أن تمــدَّ المنظمة بمعلوماتٍ مفيدة في صُنع القرار، وما إلى ذلــك. وقد انتهز العديدُ من الشركات الفرصة لتطويــر آليات جمع وتكامل وتوزيع البيانات المتخصِّصــة. وتركِّــز هذه الشركات عادةً عــلى قطاع صناعي مُحدَّد وتســتندُ إلى علاقاتها القائمة في هــذه الصناعــة من خلال منصاتها المتخصِّصــة وخدماتها لجمع البيانــات. فمثلاً؛ توفر شركة Nielsen مصــادر بيانات لعملائها حول ســلوك الــشراء بالتجزئة للعملاء. مثــالٌ آخر هو شركة Experian، والتي تجمع بيانات عن كلِّ أسرة في الولايات المتحدة. وقد طوَّرت شركة Ominiture تقنيةً لجمع نقرات الويب ومشــاركة هذه البيانات مع عملائها. وتُعَدُّ شركة Comscore إحدى كــبرى الــشركات في هذا المجال. وتقوم Google بتجميع بيانــات لمواقع الإنترنت الفردية وتقوم بعمــل مُلخص متاح من خــلال خدمات تحليلات جوجل. ومن الأمثلة الأخرى شركات Equifax، وTransUnion، وAcxiom، وMerkle، وEpsilon، وAvention. ويمكــن أن يشــملَ ذلك أيضاً منظــمات مثل ESRI.org، والتــي توفر بيانات مُوجَّهة بمواقع عملائهــا. وهناك مئات الشركات الأخرى التــي تقوم بتطوير مِنصات وخدمات متخصصة لجمع وتكامل ومشــاركة هذه البيانات مــع عملائها. وكما ذكرنا ســابقًا، يُوجَد العديدُ من دامجي ومُوزِّعــي البيانات الخاصة بالصناعة، وهــم دائمو الحركة لعرض خدمات التحليل الخاصة بهــم. وبالتالي؛ يُعَدُّ هذا القطاعُ أيضاً متزايد الاستخدام، ومورِدًا محتملًا لموهبة التحليلات، وخصوصًا مع خبرة مُتخصِّصة مُحدَّدة.

مُطوِّرو البرامج التي تركِّز على التحليلات:

قامت الشركاتُ في هذه الفئة بتطوير برامج التحليلات للاســتخدام العام مع البيانات التي تمَّ جمعها في مســتودعات البيانات أو المتاحة من خلال إحدى المنصات الأساسية المُحدَّدة سابقًا (بما في ذلــك البيانات الضخمة). كما يمكِن أن تشــمل المخترعين والباحثــين في الجامعات وغيرها من المنظمات التي طوَّرت الخوارزميات وابتكار أنواع مُحدَّدة من تطبيقات التحليلات. ويمكننا تحديدُ كبار اللاعبين في هذا المجال باستخدام أنواع التحليلات الثلاثة: الوصفية، والتنبؤية، والتوجيهية.

- **إعداد التقارير/ التحليلات الوصفية:** لقد تسَنَّى القيامُ بإعداد التقارير أو التحليلات الوصفية من خلال الأدوات التي وفَّرها مشـغلو البرمجيات الوسـيطة المشار إليها سابقًا، أو الإمكانات الفريـدة التـي يُقدِّمها مُقدِّمو الخدمـات المعنيون. فمثلاً؛ تتضَمَّـن مجموعةُ أدوات SQL Server BI التابعـة لشركة مايكروسـوفت (Microsoft) إمكانيات إعداد التقارير فضلاً عن التحليـلات التنبؤية. ومن ناحيةٍ أخرى؛ يتوفر برنامجٌ متخصصٌ من شركاتٍ أخرى مثل تابلوه (Tableau) لتصويـر البيانات. كما تُقدِّم SAS أداة التحليلات المرئية ذات السـعة نفسـها. وهناك العديدُ من أدوات التصوير المجانية أيضاً. وقد تمَّ تطويرُ مئات أدوات تصوير البيانات في جميـع أنحاء العالم، وتركِّز العديدُ من هذه الأدوات على رؤية البيانات المتوفرة لصناعة أو مجال معين. ولأن التصويرَ هو الطريق الأسـاسي حتى الآن لاستكشـاف التحليلات في مجال الصناعة؛ فقد شَـهِدَ هذا القطاع نموًا أكبر. ويتمُّ حاليًا تأسيسُ العديد من الشركات الجديدة. فمثلاً يركِّز برنامج Gephi المجاني والمفتوح المصدر على تصوير الشـبكات. وسـوف يظهر في البحث على Google أحدث قائمة لمقدِّمي مثل هذه البرامج والأدوات.
- **التحليلات التنبؤية:** ولعلَّ هذه الفئة قد شَهِدَت أكبر نموٍّ في العصر الحديث في مجال التحليلات، وهنـاك عـددٌ كبيرٌ من الـشركات التي تركِّز على التحليـلات التنبؤية. وقد تبَنَّـت العديدُ من شركات البرمجيات الإحصائية، مثل SAS وSPSS، التحليلات التنبؤية في وقتٍ مبكر، كما طوَّرت إمكانيات البرامج، وكذلك ممارساتُ الصناعة لاستخدام أساليب التنقيب في البيانات والأساليب الإحصائية الكلاسـيكية للتحليلات. ومن أمثلة الأدوات المُسـتَخدَمة للتحليلات التنبؤية IBM-SPSS Modeler من IBM وEnterprise Miner من SAS. ويشمل اللاعبون الآخرون في هذا المجال شركة KXEN وStatsoft (التي اشـترتها Dell مؤخـرًا) وSalford Systems، وكثيرٌ من الشركات الأخرى التي قد تبيع برامجها على نطاق واسع أو تستخدمها في ممارسات الاستشارات الخاصة بها (المجموعة التالية من الشركات).

كما ظهرت ثلاثُ منصات مفتوحة المصدر أخرى، وهي (R وRapidMiner وKNIME) كأدوات برمجية شائعة في المجال الصناعي للتحليلات التنبؤية، ولديها شركاتٌ تدعم تدريب وتنفيذ هذه الأدوات المجانيـة. مثل شركـة Revolution Analytics والتي تركِّز عـلى تطوير R والتدريب. ويكون تكاملُ R ممكنًا مع معظم برامج التحليلات. وتستخدم شركة تدعى Alteryx امتدادات R لإعداد التقارير والتحليلات التنبؤية، غير أن قوتها تكمُنُ في تقديم عمليات حلول التحليلات إلى العملاء والمستخدمين الآخرين بشكلٍ مشترك. وبالمثل؛ فإن RapidMiner وKNIME هما أيضاً أمثلة لموزِّدي البرامج مفتوحة المصدر. كما أن شركات مثل Rulequest التي تبيع متغيرات الملكية

مــن برامج Decision Tree وNeuroDimensions، وهي شركة برمجيات Neural Network. هي أمثلةٌ على الشركات التي طوَّرت برامج مُتخصِّصة حول أساليب مُحدَّدة للتنقيب في البيانات.

- **التحليلات التوجيهية:** يُقدِّم مُورِّدو البرمجيات في هذه الفئة أدواتِ النمذجة والخوارزميات؛ من أجل تحسينٍ أمثل للعمليات التي عادةً ما تُسمَّى برمجيات علم الإدارة/ بحوث العمليات. وكان لهذا المجال مجموعةٌ خاصةٌ من كبار مورِّدي البرمجيات. فعلى ســبيل المثال، لدى IBM برامج البرمجة الخطية الكلاســيكية، والمختلطة، والبرمجة الصحيحة. وقبل عِدة ســنوات، استحوذت IBM أيضــاً على شركة تُدعَــى ILOG، والتي تقــدِّم خدماتِ التحليــل التوجيهي والخدمات التكميليــة لتُكمِّــل عروضها الأخــرى. ومُقدِّمو التحليلات مثــل SAS لديهم أدوات OR/MS الخاصة بهم - SAS/OR. واســتحوذت شركة FICO على شركةٍ أخرى تُسمَّى XPRESS توفِّر برنامجًا للتحسين. وتشمل الجهات الفاعلة الرئيسة الأخرى في هذا المجال شركات مثل AIIMS، وAMPL، وFrontline وGAMS، وGurobi، وLindo Systems، وMaximal، وNGData، وAyata وغيرها الكثير. إن التحديد والوصف التفصيلي لعروض هذه الشركات هو خارج نطاق أهدافنا هنا. ويكفي القول بأنَّ هذا القطاع الصناعي قد شَهِدَ نموًّا كبيرًا في الآونة الأخيرة.

وبالطبــع؛ يُوجَد العديدُ من الأســاليب التــي تندرج تحت فئةِ التحليــلات التوجيهية. ولكلٍّ منها مجموعةٌ خاصةٌ بها من مُقدِّمي الخدمات. فعلى ســبيل المثــال؛ يتمُّ تقديمُ برامج المحاكاة مــن قِبل الشركات الكبرى مثــل (Rockwell (ARENA وSlink، ويوفر Palisade الأدوات التي تتضمَّن العديدَ من فئات البرامج. وبالمثل، تقدِّم Frontline أدواتٍ للتحســين باستخدام جداول بيانــات Excel؛ بالإضافة إلى التحليلات التنبؤية. ويمكِن تنفيذُ تحليل القرار في إعدادات متعددة الأغراض باستخدام أدوات مثل Expert Choice. وهناك أيضاً أدوات من شركات مثل Exsys و XpertRule وغيرها لإنشاء قواعد بشكلٍ مباشر من البيانات أو مدخلات الخبراء.

وتتطــوَّر بعــضُ الشركات الجديــدة لتجمع بين نمــاذج التحليلات المتعددة في مجــال البيانات الضخمة، بما في ذلك تحليل الشــبكات الاجتماعية، والتنقيب في تيار البيانات. فعلى ســبيل المثال، تضمــن Teradata Aster إمكانيــات التحليــلات التنبؤية التابعة له في معالجــة تدفقات البيانات الضخمة. وقد قامت العديدُ من الشركات بتطوير محرِّكات معقدة لمعالجة الأحداث (CEP) تتخذ قرارات باستخدام بيانات التدفق، مثل IBM Infosphere Streams ومايكروسوفت (Microsoft) وStreamInsight وOracle Event Processor. أمَّا الشركاتُ الكبرى الأخرى التي تمتلك منتجات محــركات معالجــة الأحــداث المعقــدة؛ فتشــمل Apache، وTibco، وInformatica، وSAP، و Hitachi. وتجدرُ الإشارة مرةً أخرى إلى أنَّ مجموعات المورِّد لجميع فئات التحليلات الثلاث ليست حصريةً. وفي معظم الحالات؛ يمكِن للمورد أن يلعب في مكونات متعددة من التحليلات.

ونقـدِّم بعد ذلـك «البتلات الداخلية» لزهـرة التحليلات. ويمكِن تسـمية هذه المجموعات مسرعات التحليلات. وعلى الرغم من أنهم قد لا يشـاركون في تطوير التقنية مباشرةً؛ فقد لعبت هذه المنظمات دورًا رئيسًا في تشكيل هذه الصناعة.

مُطوِّرو التطبيق: صناعة مُحدَّدة أو عامة:

تسـتخدم المنظماتُ في هذه المجموعة معرفتها الصناعية، وخبرتها التحليلية، وحلولها المتاحة للبنية التحتية للبيانات، ومسـتودعات البيانات، والبرامج الوسيطة ومجمِّعي البيانات ومُقدِّمي برامج التحليل الآلي لتطوير حلول مُخصَّصة لصناعة معينة. وبالتالي؛ تتيح هذه المجموعة الصناعية إمكانية اسـتخدام تقنية التحليلات في صناعة معينة. وبالطبع؛ قد تُوجَد مثل هذه المجموعات أيضاً في منظمات مستخدمين مُحـدَّدة. ومـن الواضح أنَّ معظم مُورِّدي تقنية التحليلات الأساسـيين مثل IBM، وSAS، وTeradata يتعرفون على فرصة الاتصال بصناعة مُحدَّدة أو عميل مُحدَّد وتقديم خدمات استشارية تحليلية. تقوم الشركاتُ التي قدَّمت بشكلٍ تقليدي حلول بيانات/ تطبيقات لقطاعات مُحدَّدة بتطوير عروض تحليلية خاصة بالصناعة. فعلى سـبيل المثال، تقدِّم شركة Cerner حلول السـجلات الطبية الإلكترونية لمقدِّمي الخدمات الطبية، وتشـمل عروضها الآن العديدَ من تقاريـر التحليلات والمرئيات. وبالمثل، تُقدِّم شركة IBM محرك كشـف الاحتيال لصناعة التأمين الصحي، وتعمل مع شركة تأمين لاستخدام منصة تحليلات Watson الشـهيرة في مسـاعدة مُقدِّمي الخدمات الطبية وشركات التأمين في التشخيص وإدارة المرض. ومـن الأمثلة الأخرى على مُورِّدي تطبيقات رأسـية Sabre Technologies، التـي تُقدِّم حلولًا تحليليةً لصناعة السَّفر بما في ذلك تسعير التعريفات لتحسين الإيرادات وتخطيط الإرسال.

وتشـمل هذه المجموعة أيضاً الشركاتِ التي طوَّرت حلول تحليلات خاصة بمجالها وتسـوقها على نطاقٍ واسع لقاعدة العملاء. وعلى سبيل المثال، تقوم Nike، وIBM، وSportvision بتطوير تطبيقات في التحليلات الرياضية لتحسين اللعب وزيادة نسبة المشاهدة. ولقد طوَّرت Acxioni مجموعاتٍ لكلِّ المنازل في الولايات المتحدة تقريبًا اسـتنادًا إلى البيانات التي تجمعها حول الأسر من مصادر مختلفة. وتنتمي شركات إعداد التقارير ذات الدرجات الائتمانية والتصنيف (FICO، وExperian... إلـخ) أيضاً إلى هذه المجموعة. وتقدِّم IBM والعديد من الشركات الأخرى حلول تحسين الأسعار في قطاع بيع التجزئة.

ويمثـل هذا المجـال فرصةً رياديةً لتطوير تطبيقات خاصة بالقطـاع. ويحاول الكثيرون الذين يظهرون في تحليلات الويب/ الشبكات الاجتماعية/ المواقع أن يقوموا بتهيئة المستخدمين للاستهداف الأفضل للحملات الترويجية في الوقت المناسـب. ومن أمثلة هذه الشركات وأنشطتها: تقوم شركة YRcom باسـتخدام بيانات الموقع لتطوير ملفات تعريف المسـتخدمين/ المجموعات واستهداف

إعلانات الجوال، ومســتخدمي ملفات تعريف Towerdata على أساس استخدام البريد الإلكتروني. وتهدف شركة Qualia إلى تحديد المســتخدمين من خلال جميع اســتخدامات الجهاز، وتستهدف Simulmedia الإعلانات التلفزيونية على أساس تحليل عادات مشاهدة التلفزيون لدى المستخدم.

ولقد أنتج نموُّ الهواتف الذكية صناعةً كاملةً تركِّز على تطبيقات تحليلات مُحدَّدة للمستهلكين وكذلك المؤسســات. فعلى سبيل المثال: يمكِن لتطبيقات الهواتف الذكية مثل Shazam، أو Soundhound، أو Musixmatch تحديد أغنية على أســاس المشــاهدات القليلة الأولى، ثم السماح للمستخدم باختيارها بغرض التشغيل/ التنزيل/ الشراء. وتستخدم Waze معلومات المرور الفورية التي يشاركها المستخدمون. بالإضافة إلى بيانات الموقع لتحسين التنقل. وتؤدي أدوات التعرُّف على الصوت مثل Siri على iPhone وGoogle Now وAmazon Alexa إلى العديــد مــن تطبيقــات التحليــلات الأكــثر تخصصًــا لأغــراض مُحــدَّدة جــدًا في التحليــلات المطبَّقة عــلى الصــور ومقاطع الفيديــو والصــوت والبيانات الأخــرى التــي يمكِــن التقاطها من خــلال الهواتــف الذكيــة و/أو أجهزة الاستشــعار المتصلــة. لقد رفعــت الهواتــف الذكيــة أيضــاً مــن مــورِّدي الخدمــات الاقتصاديــة المشــتركة، مثــل: Uber، وLyft، وCurb، وOla. إنَّ العديــد مــن هــذه الشركات؛ هــي أمثلةٌ للتحليلات التــي تؤدي إلى فرص أعــمال جديدة. وتُعَدُّ وســائلُ التواصل الاجتماعي عبر الإنترنت مجالًا آخر ســاخنًا في هذه المجموعة. ومما لا شــك فيه أن Facebook؛ هو اللاعب الأســاسي في مجال الشبكات الاجتماعية عبر الإنترنت يليه Twitter وLinkedln. وعــلاوةً عــلى ذلك؛ فإنَّ الوصول العام إلى بياناتهم أدَّى إلى ظهور شركاتٍ أخرى متعددة تقوم بتحليل بياناتها. فعلى ســبيل المثال: تقــوم Unmetric بتحليل بيانات Twitter، وتوفير الحلول لعملائها، وبالمثل، يُوجَد العديدُ من الشركات الأخرى التي تركِّز على تحليل الشبكات الاجتماعية.

ومن المجالات الموجَّهة في مجال تطوير التطبيقات إنترنت الاشياء IoT. وتقومُ العديدُ من الــشركات ببناء تطبيقات لصنع كائنات ذكية. فمثلًا، طوَّرت شركة SmartBin أنظمة المراقبة الذكية عن بُعد لقطاعات النفايات وإعادة التدوير، وتعمل العديدُ من المنظمات الأخرى على بناء العدَّادات الذكية والشــبكات الذكية والمدن الذكية والســيارات المتصلة والبيوت الذكية، وسلاسل التوريد الذكية، والصحة المتصلة والتجزئة الذكية، وغيرها من الأشياء الذكية.

وينمو هذا النشــاط في بداية الأمر ليتحوَّل إلى مرحلة انتقالية كبيرة بسبب مشكلات في التمويل/ الاستثمار والمشكلات الأمنية/ الخصوصية. وعلى الرغم من ذلك؛ فقد يكون قطاع مطوِّري التطبيقات هــو أكبرُ مجالٍ للنمو داخل التحليلات في هــذه المرحلة. وتوفر هذه المجموعة فرصةً فريدةً لمحترفي التحليلات الذين يبحثون عن المزيد من خيارات مهنة تنظيم المشاريع.

مُحلِّلو صناعة التحليلات والمؤثرون فيها:

وتشــمل المجموعــة التالية من صناعــة التحليلات ثلاثــةَ أنواعٍ من المنظــمات أو المحترفين؛ المجموعــة الأولى هــي مجموعةٌ مــن المنظمات المهنية التي تُقدِّم المشــورة إلى مــورِّدي صناعة التحليلات والمستخدمين. تشمل خدماتهم تحليلات تسويقية وتغطية التطورات الجديدة وتقييم تقنيــات مُحدَّدة وتطوير التدريــب/ الورقات البيضاء وما إلى ذلك. ومــن أمثلة هذه المجموعة منظــمات مثــل مجموعة جارتنر، ومعهــد مســتودعات البيانــات، وForrester، وMcKinsey. والعديد من المنشــورات العامــة والفنية ومواقع الويب التي تغطــي صناعة التحليلات. وتتميز مجموعة ماجيك كوادرانتــس التابعة لمجموعة جارتنر بكونها مؤثرةً للغاية وتعتمد على عمليات المســح في الصناعة. وبالمثل، يوفِّر محترفو TDWI.org لمحــاتٍ عامة ممتازة عن الصناعة، وهم يدركون تمامًا الاتجاهات الحالية والمستقبلية لهذه الصناعة.

وتضــمُّ المجموعة الثانية مجتمعات أو شركات مهنية تقدِّم أيضاً بعضًا من الخدمات نفســها؛ ولكنها قائمةٌ على أساسٍ منظم. فعلى سبيل المثال: يركِّز الآن معهدُ بحوث العمليات وعلم الإدارة INFORMS، وهي شركةٌ مهنيــةٌ، على تعزيز التحليلات. وتركِّز مجموعة المصالح الخاصة لدعم القرار والتحليلات، وهي مجموعةٌ فرعيةٌ تابعةٌ لرابطة نظم المعلومات، أيضاً على التحليلات. كما أنَّ لدى معظم المورِّدين الرئيســيين (مثل: تيراداتا، وSAS) مجموعات المســتخدمين الخاصة بهم. وتشــجِّع هذه الكيانات اســتخدامَ التحليلات وتمكِّن من مشــاركة الدروس المستفادة من خلال منشوراتها ومؤتمراتها، وقد توفر أيضاً خدمات التوظيف، وبالتالي، مصادر جيدة لتحديد المواهب.

أمَّــا المجموعة الثالثة من المحلِّلين في مجال التحليلات؛ فهي ما نســميه ســفراء التحليلات، أو المؤثرين، أو القائمين بالدعاية. وقد أبدى هؤلاء المحلِّلون حماســهم للتحليلات من خلال حلقاتهم الدراســية، وكتبهم ومنشــورات أخرى، وتشمل هذه المجموعة على ســبيل المثال لا الحصر Steve Baker، وTom Davenport، وCharles Duhigg، وWayne Eckerson، وbill Franks، و Malcolm Gladwell، وClaudia Imhoff، وbill Inman، وغيرهــم. ولدى كلٍّ هؤلاء الســفراء مؤلفات (بعضها من الكتب الأكثر مبيعًا) و/أو قدَّموا العديدَ من العروض التقديمية لترويج تطبيقات التحليلات. وقد يكون هناك مجموعةٌ أخرى من المؤثرين لذكرها هنا، وهم مؤلفو الكتب المدرسية عــن تحليلات ذكاء الأعمال الذين يهدفون إلى مســاعدة المجموعة التاليــة لإنتاج محترفي صناعة التحليلات. ومن الواضح أنَّ الأمر سيســتغرق بعض الوقت؛ لكي يصبح طالبُ التحليلات عضوًا في هذه المجموعة؛ ولكنه يستطيع العمل مع أعضاء من هذه المجموعة كالباحثين أو المتدربين.

المعاهد الأكاديمية ووكالات الاعتماد:

تستمدُّ أيُّ صناعةٍ كثيفةِ المعرفة، مثل التحليلات؛ قوتَها الأساسية من جذبِ الطلاب المهتمين بالتقنية واختيار هذه الصناعة كمهنةٍ لهم. وتلعبُ الجامعات دورًا رئيسًا في جَعْل ذلك ممكنًا. وتُقدِّم هذه المجموعة، البرامج الأكاديمية التي تُعِدُّ المهنيين لهذه الصناعة. وهي تشملُ مكوناتٍ مختلفةً من كليات إدارة الأعمال، مثل نظم المعلومات والتسويق وعلوم الإدارة، وما إلى ذلك. كما يمتدُّ إلى ما هو أبعد من المدارس التجارية لتشمل أقسام علوم الحاسب والإحصاء والرياضيات والهندسة الصناعية في جميع أنحاء العالم. كما تمتدُّ كذلك لتشمل ما يتجاوز مُطوِّري الرسومات الذين يُصمِّمون طرقًا جديدةً لتصوير المعلومات. وتقدِّم الجامعات برامج البكالوريوس والدراسات العليا في التحليلات في جميع هذه التخصصات، على الرغم من اختلاف التسميات. وقد اشتملت إحدى نطاقات النموِّ الرئيسة في برامج الشهادات على التحليلات؛ لتمكين المتخصِّصين الحاليين من إعادة تدريبهم وإعادة تنظيم أنفسهم في وظائف التحليلات. وتمكِّن برامج الشهادات الممارسين من اكتساب الكفاءة الأساسية في برامج معينة عن طريق الحصول على عددٍ قليلٍ من الدورات المهمة بالمدارس التي تقدِّم هذه البرامج. وتتضمَّن شبكة جامعة تيراداتا قائمةً ببرامج التحليلات، والتي تشملُ ١٥٠ برنامجًا تقريبًا، وتتزايد يوميًّا.

وتساعدُ مجموعةٌ أخرى من اللاعبين في تطوير الكفاءة في التحليلات. وهذه هي برامج الشهادات التي تمنحُ شهادة الخبرة في برامج مُحدَّدة. ويقدِّم كلُّ مورِّدي التقنية الأساسية تقريبًا [أي بي إم (IBM)، ومايكروسوفت (Microsoft) ومايكروستراتيجي (Microstrategy)، وأوراكل (Oracle)، وإس إيه إس (SAS)، وتابلوه (Tableau) وتيراداتا (Teradata)]؛ برامج الشهادات الخاصة بها. وتضمَن هذه الشهاداتُ أن يتمتع الموظفون المحتملون الجُدد بمستوًى معين من الأدوات المهارية. ومن ناحيةٍ أخرى؛ تقدِّم INFORMS برنامج شهادة Certified Analytics Professional، والذي يهدفُ إلى اختبار الكفاءة التحليلية العامة لشخصٍ ما. أي إنَّ هذه الشهادات تمنح الطالبَ الجامعيَّ مهاراتٍ إضافيةً قابلةً للتسويق.

يُعَدُّ نموُّ البرامج الأكاديمية في التحليلات أمرًا مذهلًا. ويُحدِّد الوقت فقط ما إذا كانت هذه المجموعة تفرط في زيادة السعة التي يمكن أن تستهلكها المجموعات الأخرى. ولكن في هذه المرحلة. يبدو أن الطلبَ يفوق المعروض من خريجي التحليلات المؤهلين، وهذا هو المكانُ الأكثر وضوحًا للعثور على مستخدمي التحليلات المبتدئين على الأقل.

المنظمون وصُنَّاع السياسات:

يُعتَبَر اللاعبون في هذا المكوِّن مسؤولين عن تحديد القواعد واللوائح الخاصة بحماية موظفي، وعمــلاء، وحاملي أســهم منظمات التحليلات. ويتطلَّب جمعُ ومشــاركةُ بيانــات العميل قوانينَ صارمةً لتأمين الخصوصية. تقوم عِدة منظمات في هذا المجال بتنظيم عملية نقل البيانات وحماية حقوق المســتخدمين. فعلى سبيل المثال: تُنظِّم لجنة الاتصالات الفيدرالية (FCC) الاتصالات بين الــدول والاتصــالات الدولية. وبالمثل. فــإن لجنة التجارة الفيدرالية (FTC) هي المســؤولة عن عرض الممارســات التجارية غير المشروعــة المتعلقة بالبيانات. ويُنظِّــم الاتحاد الدولي للاتصالات (ITU) الوصــول إلى تقنية المعلومات والاتصــالات (ICTs) للمجتمعات غير المخدومة في جميع أنحاء العالم. ومن ناحيةٍ أخرى؛ يُوجَد وكالةٌ فيدراليةٌ غير نظامية تُسمَّى المعهدُ الوطني للمعايير والتقنية (NIST)، تســاعد على تقدُّم البنية التحتية للتقنية. كما أنَّ هناك العديدَ من المنظمات الأخــرى في جميع أنحاء العالم التــي تنظم أمن البيانات وتدفع عجلــة صناعة التحليلات. وهذا مكونٌ مهمٌ في النظام البيئي، وبالتالي لا يمكِن لأحدٍ أن يُسيء استخدامَ معلومات المستهلكين.

وبالنســبة لأيِّ شخصٍ يقومُ بتطوير أو استخدام تطبيقات التحليلات، ربما يكون من الضروري أن يكون هناك شــخصٌ ما في الفريق على علمٍ بالإطار التنظيمي. ومن الواضح أنَّ هذه الوكالات والمنظمات المهنية التي تعملُ معهم يقدِّمون مواهبَ ومهاراتٍ فريدةً في مجال التحليلات.

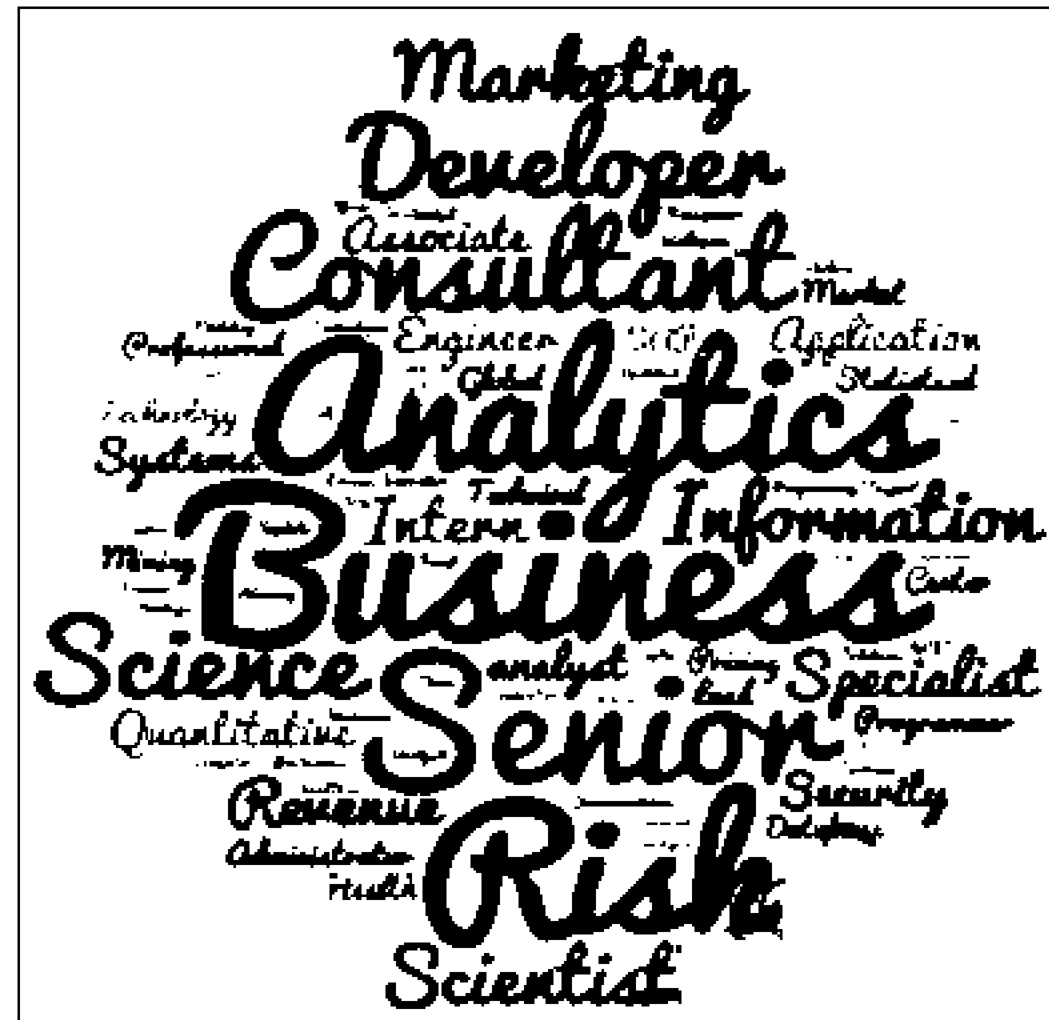
منظمات مُستخدِمي التحليلات:

مــن الواضح أنَّ هذا هو المحركُ الاقتصادي لصناعة التحليــلات بأكملها، وبالتالي؛ فإننا نمثِّل هذه المجموعــة باعتبارهــا جوهرَ زهــرة التحليلات. وإذا لم يكن هناك مســتخدمون؛ فلــن تكون هناك صناعةُ تحليلات. إن المنظمات في كلِّ صناعة، بغض النظر عن حجمها وشــكلها وموقعها؛ تســتخدمُ أو تستكشــفُ استخدام التحليلات في عملياتها. وتشــمل هذه المنظمات القطاع الخاص، والحكومة، والتعليم، والعســكرية، وما شابهها؛ بل وتشــمل المنظمات في جميع أنحاء العالم. وهناك أمثلةٌ على استخدامات التحليلات في صناعات مختلفة كثيرة. ويستكشف آخرون فرصًا مماثلةً لمحاولة الحصول على مَيْزة تنافســية أو الاحتفاظ بها. ولم يتم تحديدُ شركات مُحدَّدة في هذا القســم؛ وبدلًا من ذلك؛ فإن الهدف هنا هو معرفةُ نوع أدوار محترفي التحليلات التي يمكن أن يلعبها داخل منظمة المُستخدِم.

وبالتاكيد؛ تُعَدُّ القيادةُ العليا للمنظمة لا سيما في مجموعة تقنية المعلومات (مسؤول المعلومات، وما إلى ذلك)؛ أمرًا بالغ الأهمية في تطبيق التحليلات لعملياتها. ويقول Forrest Mars من إمبراطورية

Mars للشوكولاتة: «إنَّ جميعَ الإدارات تُغالي في تطبيق الرياضيات على عمليات واقتصاديات الشركة». وعلى الرغم من عدم تأييد عددٍ كافٍ من كبار المديرين لهذا الرأي؛ فإنَّ الوعي بتطبيق التحليلات داخل المنظمة ينمو في كلِّ مكان. ولقد أخبرنا مسؤول تنفيذي في شركة تأمين صحي ذات مرة أنَّ (المدير التنفيذي) ينظر إلى الشركة كمنظمةٍ تمكِّن تقنية المعلومات من جمع الأموال من الأعضاء المؤمَّن عليهم وتوزعها على المورِّدين. وهكذا كانت الكفاءة في هذه العملية؛ هي الخاصية التي يمكن أن يتفوقوا بها على المنافس. وقد أدَّى هذا إلى تطوير العديد من تطبيقات التحليلات لتقليل الاحتيال والمدفوعات الزائدة لمقدِّمي الخدمات، وتعزيز العافية بين المؤمَّن عليهم؛ حتى يستخدموا مورِّدي الخدمات بمعدلٍ أقل من المعتاد، وتوليد المزيد من الكفاءة في المعالجة، وبالتالي تكون أكثر ربحيةً.

ومن الناحية الفعلية؛ فإنَّ جميع المنظمات الرئيسة في كلِّ صناعةٍ محل الاهتمام؛ توظِّف محترفي التحليلات تحت مُسميات وظيفية مختلفة. ويُعَدُّ الشكل ١-١٤ تصويرًا لكلمة واحدة من عناوين مختارة من خريجي برنامجنا في جامعة أوكلاهوما الحكومية من عام ٢٠١٣ إلى عام ٢٠١٦. ويُبيِّن بوضوح أنَّ تحليلات وعلم البيانات هي عناوين شائعة في المنظمات التي توظف خريجي هذه البرامج. وتظهر الكلمات الرئيسة الأخرى لتتضمَّن مصطلحات مثل المخاطر وقواعد البيانات والأمن. والإيرادات والتسويق، وما إلى ذلك.



شكل ١-١٤: سحابة الكلمات للألقاب الوظيفية لخريجي برنامج التحليلات

وبالطبع؛ تتضمَّن منظماتُ المســتخدمين مساراتٍ مهنيةً لمحترفي التحليلات للانتقال إلى مواقع الإدارة. وتشمل هذه المسميات الوظيفية مديري المشروعات، وكبار المديرين، والقادة، وصولًا إلى رئيس قســم المعلومات أو الرئيس التنفيذي. وهذا يشــيرُ إلى أنَّ منظمات المستخدمين موجودةٌ كمجموعة رئيسة في النظام البيئي للتحليل، وبالتالي يمكِن أن يكون مصدرًا جيدًا للمواهب. وربما يكون هذا هو المكان الأول لإيجاد محترفي التحليلات في قطاع الصناعة الرأسية.

وكان الغرضُ من هذا القســم هو تقديمُ لمحةٍ سريعة عن مَشْــهد صناعة التحليلات. وقد تمَّ تحديــدُ ١١ مجموعةً مختلفةً تلعبُ دورًا رئيسًــا في بناء هــذه الصناعة وتعزيزها. ويمكِن إضافة المزيد من البتلات/ المكونات بمرور الوقت في زهرة/ نظام التحليلات. ونظرًا لأن تحليلات البيانات تتطلَّــب مجموعــةً متنوعةً من المهارات؛ فإن فهمَ هذا النظام البيئــي يوفِّر لك خيارات أكثر مما قــد تتخيل لمهن التحليلات. وعلاوةً على ذلك؛ يمكِن للمحترفين الانتقالُ من مجموعةٍ صناعية إلى أخرى للاستفادة من مهاراتها. فعلى سبيل المثال: يمكِن للخبراء المحترفين من مُقدِّمي الخدمات في بعض الأحيان الانتقالُ إلى مراكز الاستشــارات، أو مباشرةً إلى منظمات المستخدمين. وبشكلٍ عام؛ هناك الكثيرُ مما يثيرُ الحماسة حول صناعة التحليلات في هذه المرحلة.

أسئلة مراجعة على القسم ١-٨:

١- قم بإدراج ١١ فئةً من اللاعبين في النظام البيئي للتحليل.

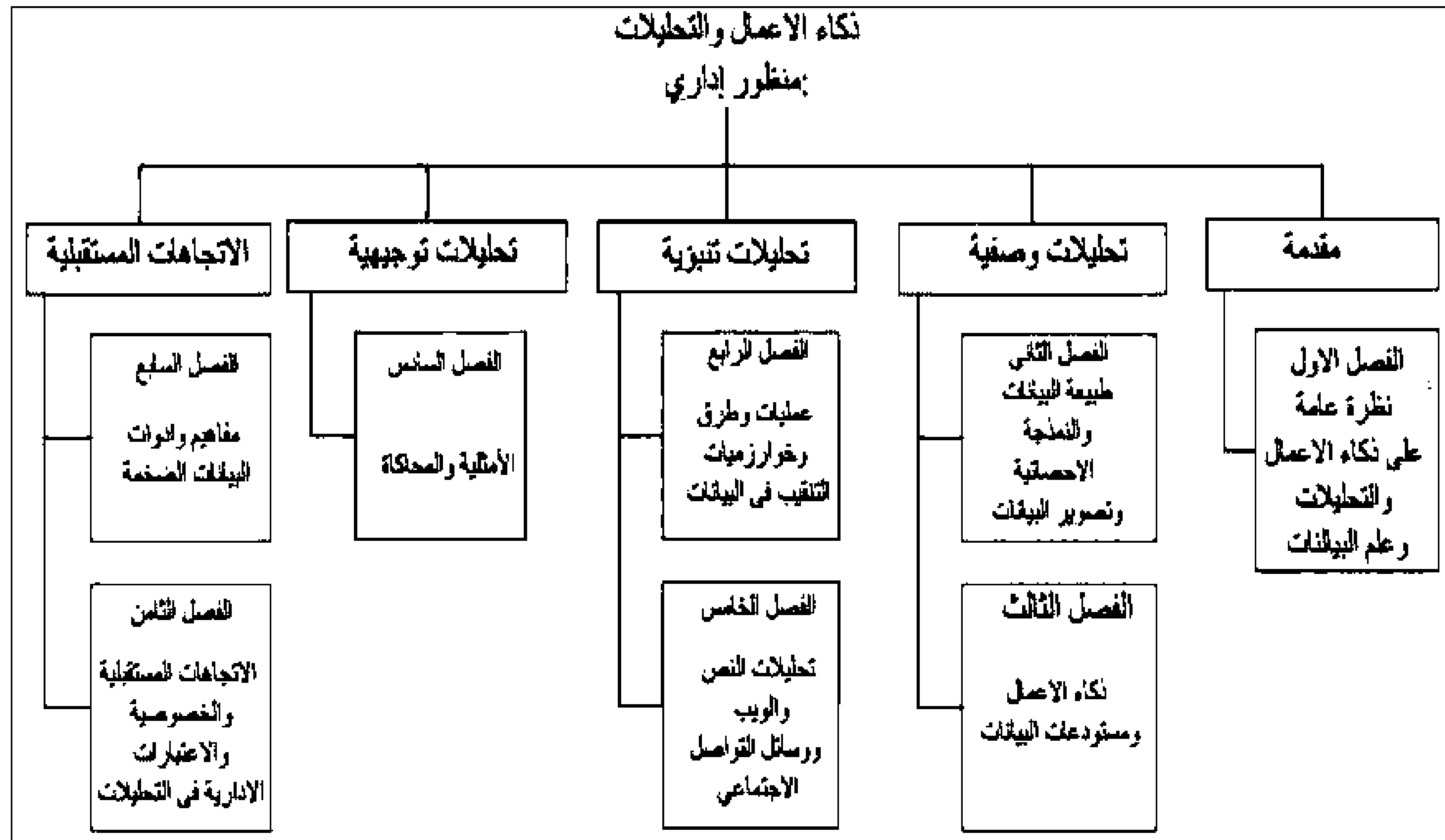
٢- أعطِ أمثلةً لشركات في كل واحد من الـ ١١ نوعاً من اللاعبين.

٣- ما هي الشركات المهيمنة في أكثر من فئة؟

٤- هل من الأفضلِ أن تكونَ أقوى لاعبٍ في فئة واحدة أو أن تكون نشطًا في فئاتٍ مُتعدِّدة؟

١-٩ خطة الكتاب:

لقد منحتكَ الأقسامُ السابقةُ فهمًا لحاجة تقنية المعلومات في صُنع القرار، وتطوُّر ذكاء الأعمال، والآن إلى التحليلات وعلم البيانات. وقد قدَّمنا في الأقسام العديدة الأخيرة نظرةً عامةً على أنواع مختلفة من التحليلات وتطبيقاتها. والآن نحن مســتعدون للقيام برحلةٍ إرشــادية أكثر تفصيلًا في هــذه المواضيع، إلى جانب بعض الخــبرة العملية العميقة في بعض الموضوعــات الفنية، ويوضِّح الشكل ١-١٥ خطةً حول بقية الكتاب.



شكل ١-١٥: خطة الكتاب

في هذا الفصل؛ تمَّ عرضُ مقدمة وتعريفات ونظرة عامة على أنظمة دعم القرار DSSs وذكاء الأعمال والتحليلات، بما في ذلك تحليلات البيانات الضخمة وعلم البيانات. كما قدَّمنا نظرةً عامةً على النظام البيئي للتحليلات لتقدِّر مدى اتســاع وعمق هــذه الصناعة. ويُغطِّي الفصلان الثاني والثالث المسائل التحليلية الوصفية والبيانات. وتشكِّل البيانات غالبًا الأساس لأيِّ تطبيق تحليلي. وبالتالي نقومُ بعرض مقدمة حول مســائل مستودعات البيانات والتطبيقات والتقنيات. ويغطي هذا القسمُ أيضاً تقنيات وتطبيقات التقارير وتطبيقات الأعمال. ويتبع ذلك نظرة عامة مختصرة على أساليب وتطبيقات إدارة أداء الأعمال BPM، وهو الموضوع الذي كان جزءًا أساسيًّا من ذكاء الأعمال التقليدي.

ويُغطِّي القســمُ التالي التحليلات التنبؤية؛ إذ يُقدِّم الفصل الرابع مقدمةً لتطبيقات التحليلات التنبؤية، ويتضمَّن العديدَ من الفنيات الشــائعة للتنقيب في البيانات وهي: التصنيف والتجميع وأســاليب الترابط، وغيرها. ويركِّز الفصلُ الخامس على تطبيقات التنقييب في النصِّ بالإضافة إلى تحليــلات الويــب، بما في ذلك تحليلات وســائل التواصل الاجتماعي وتحليلات المشــاعر. وغيرها من الموضوعات ذات الصِّلة. ويُغطِّي الفصل الســادس التحليــلات التوجيهية. ويتضمَّن الفصل الســابع مزيدًا من التفاصيل حول تحليلات البيانات الضخمة. ويتضمَّن الفصلُ الثامن مناقشــة الاتجاهــات الحديثــة الظهور. مما يؤدي إلى انتشــار كلٍّ من الأجهزة اللاســلكية، وأجهزة GPS،

وأجهزة الاستشــعار الأخرى وإنشــاء قواعد بيانات ضخمة جديــدة وتطبيقات فريدة من نوعها. وقد بــدأت مجموعةٌ جديدةٌ من شركات التحليلات في الظهــور لتحليل قواعد البيانات الجديدة هذه؛ من أجل فهمٍ أفضل وأكثر عمقًا لســلوكيات وتصرُّفات العملاء، وهو ما يؤدي إلى وَضْع آلية للتحليلات والتوسُّــع إلى مجال جديد يُســمَّى إنترنت الأشياء IoT ويغطِّي الفصلُ أيضاً التحليلات المســتندة إلى الســحابة. وأخيرًا؛ يحاول الفصلُ الثامن أيضاً دمجَ جميــع المواد التي يغطيها هذا الكتاب، ويُختَتم بمناقشةٍ موجزةٍ لأبعاد الأمان/ الخصوصية في التحليلات.

١-١٠ موارد، وروابط، واتصال شبكة جامعة تيراداتا:

يمكِن تعزيزُ اســتخدام هذا الفصل ومعظم الفصول الأخرى في هذا الكتاب من خلال الأدوات الموضَّحة في الأقسام التالية.

الموارد والروابط:

نوصِي بالموارد والروابط الرئيسة التالية:

- The Data Warehousing Institute (tdwi.org)
- Data Science Central (datasciencecentral.com)
- DSS Resources (dssresources.com)
- Microsoft Enterprise Consortium (enterprise.waltoncollege.uark.edu/mec.asp)

الموِّردون والمنتجات والنسخ التجريبية:

يُقدِّم معظمُ الموِّردين نسخًا تجريبيةً لمنتجاتهم وتطبيقاتهم. وتتوافر معلوماتٌ حول المنتجات والبنية والبرمجيات على dssresources.com.

النشرات الدورية:

نوصِي بالنشرات الدورية التالية:

- Decision Support Systems (www.journals.elsevier.com/decision-support-systems)
- CIO Insight (cioinsight.com)

اتصال شبكة جامعة تيراداتا:

يرتبــط هــذا الكتــاب ارتباطًا وثيقًا بالمــوارد المجانيــة التي تقدِّمها شــبكةُ جامعــة تيراداتا (انظر: teradatauniversitynetwork.com). وتنقسم بوابة شبكة جامعة تيراداتا إلى قسمين رئيسين: أحدهما للطلاب، والآخر لهيئة التدريس. وهذا الكتابُ مرتبطٌ ببوابة شــبكة جامعة تيراداتا عبر قســمٍ خاصٍّ في نهاية كلِّ فصل. ويتضمَّن هذا القسمُ روابط مناسبةً للفصل المحَدَّد. ويشيرُ إلى الموارد ذات الصِّلة. كما نقدِّم تدريبًا عمليًّا باستخدام البرمجيات وغيرها من المواد (كالحالات) المتاحة في شبكة جامعة تيراداتا.

موقع الكتاب على الإنترنت:

موقــع هذا الكتاب على الإنترنت هو: pearsonhighered.com/sharda، ويحتوي على نصٍّ إضــافيٍّ تكميليٍّ منظم كفصول ويب تتوافق مع فصول الكتاب المطبوع. ويتمُّ سَرْد مواضيع هذه الفصول في جدول محتويات الفصل عبر الإنترنت.

ويُرجَــى ملاحظــة أنه في أثناء مثولِ هذا الكتاب للطباعة؛ تحققنا من أن جميع مواقع الويب التي تمَّ الاستشــهادُ بها كانت نشــطةً وصالحة. وعلى أيِّ حال؛ تكون عناوين URL ديناميكية. كما تتغيرُ أحيانًا مواقع الويب التي نشيرُ إليها في النص أو تتوقف بسبب تغيير الشركات لأسمائها أو شرائهــا أو بيعها أو دمجها أو تعطلها. وفي بعض الأحيان تكون مواقع الويب تحت الصيانة أو الإصلاح أو إعادة التصميم. وقد أســقطت العديدُ من المؤسســات البادئة "www" من مواقعها؛ ولكن البعض الآخر لا يزال يســتخدمها. وإذا واجهتك أيُّ مشــكلة في الاتصال بموقع الويب الذي نذكره؛ فيُرجَى التحلِّي بالصبر وببساطة قم بعمل بحث على الويب؛ لمحاولة تحديد الموقع الجديد المحتمَل. والذي في الغالب سيمكنك العثور عليه بسرعة من خلال أحد محركات البحث الشائعة. ولا يسعنا إلا أن نعتذر مقدمًا عن هذا الإزعاج.

تلخيص لأهمِّ نقاط الفصل:

- أصبحت بيئة الأعمال أكثرَ تعقيدًا وتتغير بسرعة؛ مما يجعل اتخاذ القرار أكثر صعوبةً.
- يجب أن تستجيب الشركاتُ، وتتكيف مع بيئة التغيير بسرعة عن طريق اتخاذ قرارات أسرع وأفضل.
- إن الإطــار الزَّمني لاتخاذ القــرارات يتقلَّص، في حين أن الطبيعة العالميــة لصُنع القرار آخذة في التوسع؛ مما يستلزم تطوير واستخدام نظم إدارة أمن البيانات المحوسبة.
- تســتخدم أنظمةُ دعم القرار البيانات والنماذج وأحيانًا إدارة المعرفة لإيجاد حلول للمشــكلات شِبه المهيكلة وبعض المشكلات غير الهيكلية.

- تَستخدم طرق ذكاء الأعمال مستودعًا مركزيًا يُسمَّى مستودع البيانات، والذي يمكِّن من التنقيب في البيانات بكفاءة وتصوير البيانات.
- وتشـمل بنيةُ ذكاء الأعمال مسـتودعاتِ البيانات وأدوات تحليل الأعمال المُسـتَخدَمة من قِبل المستخدمين النهائيين وواجهة المستخدِم (مثل لوحات المعلومات).
- تَستخدم العديدُ من المنظمات التحليلات الوصفية لتحلَّ محلَّ تقاريرها التقليدية المُسطَّحة مع التقارير التفاعلية التي تقدِّم رؤى واتجاهات وأنماط في بيانات المعاملات.
- تمكِّـن التحليلاتُ التنبؤية المؤسسـات من وَضْع قواعد تنبؤية تُشـجِّع نتائج الأعمال من خلال تحليل البيانات التاريخي للسلوك الحالي للعملاء.
- تساعد التحليلاتُ التوجيهية في بناء النماذج التي تنطوي على أساليب التنبؤ والأمثلية على أساس مبادئ تقارير التشغيل وعلم الإدارة؛ لمساعدة المنظمات على اتخاذ قرارات أفضل.
- يُركِّـز تحليـل البيانـات الضخمـة عـلى مجموعـات البيانـات الكبيرة غـير المهيكلـة، والتي قد تتضمَّن أيضاً أنواعًا مختلفةً تمامًا من البيانات للتحليل.
- تُعرف التحليلات كمجالٍ بأسـماء التطبيقات الخاصة بالصناعة، مثل التحليلات الرياضية، وهو معروفٌ أيضاً بأسماء أخرى ذات صِلة مثل علم البيانات أو علم الشبكة.
- إنَّ الرِّعاية الصحية وسلاسـل بيع التجزئة؛ هما مجالان تكثر فيهما تطبيقات التحليلات، والكثير من المجالات قادمة.
- يمكـن النظـرُ إلى النظـام البيئي للتحليـل لأول مرة على أنـه مجموعةٌ مـن مُقدِّمي الخدمة والمستخدمين والميسرين. يمكن تقسيمها إلى ١١ مجموعةً.

مصطلحات أساسية:

التحليلات	لوحة المعلومات	التحليلات الوصفية (أو إعداد التقارير)
معالجة المعاملات الفورية (OLTP)	النظام البيئي للتحليلات	التنقيب في البيانات
تحليلات البيانات الضخمة	القرار أو التحليلات المعيارية	وكلاء الذكاء
التحليلات التنبؤية	ذكاء الأعمال (BI)	المعالجة التحليلية الفورية (OLAP)
التحليلات التوجيهية	خدمات الويب	

أسئلة للمناقشة:

١- قم بإجراء مَسْـح الدراسات منذ الأشهر الستة الماضية؛ للعثور على تطبيقٍ واحدٍ لكلٍّ من نُظُم دعم القرار DSS، وذكاء الأعمال والتحليلات. ولخِّص التطبيقات في صفحة واحدة، وأرسـلها مع عرضٍ دقيق.

٢- ميِّز بين ذكاء الأعمال ونظام دَعْم القرار.

٣- قارن وميِّز بين التحليلات التنبؤية والتحليلات الوصفية والتحليلات التوجيهية. استخدم الأمثلة.

٤- ناقش المشكلات الرئيسة في تطبيق ذكاء الأعمال.

تمارين شبكة جامعة تيراداتا (TUN) وغيرها من التمارين اليدوية:

١- اذهب إلى teradatauniversitynetwork.com. وباستخدام كلمة مرور الموقع التي يقدِّمها معلمك، قم بالتسجيل في الموقع إذا لم تكن قد قمتَ بالتسجيل مسبقًا. قم بتسجيل الدخول وتعرَّف على محتوى الموقع. سوف تتلقى المهمات المتعلقة بهذا الموقع. قم بإعداد قائمة تضمُّ ٢٠ عنصرًا على الموقع تعتقد أنها قد تكون مفيدةً لك.

٢- اذهب إلى موقع شـبكة جامعة تيراداتا. استكشف صفحة Sports Analytics، وقم بتلخيصِ اثنين على الأقل من تطبيقات التحليلات في أيٍّ رياضة من اختيارك.

٣- ادخل إلى موقع شبكة جامعة تيراداتا. وحَدِّد «الحالات والمشاريع والواجبات». ثم اختر دراسة الحالـة «Harrah's High Payoff from Customer Information». أجب عن الأسـئلة التالية حول هذه القضية:

   ١- ما هي المعلومات التي تنتج عن التنقيب في البيانات؟
   ٢- كيف تكون هذه المعلومات مفيدةً للإدارة في صُنع القرار؟ (كن دقيقًا.)
   ٣- اذكر أنواع البيانات التي يتمُّ التنقيب فيها.
   ٤- هل هذا تطبيق DSS أو BI؟ ولماذا؟

٤- اذهـب إلى teradatauniversitynetwork.com، وقم بإيجاد الورقة بعنوان «مسـتودعات البيانـات يدعـم إسـتراتيجية الشركـة في أول شركـة أمريكيـة» ش (بواسـطة Watson وWixom وGoodhue). اقرأ الورقة، وأجب عن الأسئلة التالية:

   ١- ما هي دوافع مشروع مستودع البيانات / ذكاء الأعمال في الشركة؟

٢- ما هي المزايا الإستراتيجية التي تحققت؟

٣- ما هي المزايا التشغيلية والتكتيكية التي تحققت؟

٤- ما هي عوامل النجاح الحسَّاسة بالنسبة للتنفيذ؟

٥- انتقل إلى http://analytics-magazine.org/issues/digital-editions وقم بإيجاد إصدار يناير / فبراير ٢٠١٢م بعنوان «مشكلة خاصة: مستقبل الرعاية الصحية». اقرأ «التحليلات التنبؤية - إنقاذ الأرواح وخفض الفواتير الطبية». أجب عن الأسئلة التالية:

١- ما هي المشكلة التي يتمُّ تناولها من خلال تطبيق التحليلات التنبؤية؟

٢- ما هي درجة الالتزام بأدوية FICO؟

٣- كيف يتمُّ التدريبُ على نموذج التنبؤ بأدوية FICO؟ وهل صنَّف نموذج التنبؤ درجة التزام الدواء في FICO؟

٤- قم بتكبير الشكل ٤، واشرح أيّ نوعٍ من الأساليب يتمُّ تطبيقه على النتائج المتولدة.

٥- اذكر بعض القرارات القابلة للتطبيق التي تمَّ استخدامها بناءً على نتائج التنبؤ.

٦- انتقل إلى http://analytics-magazine.org/issues/digital-editions. وقم بإيجاد إصدار يناير/ فبراير ٢٠١٣ «العمل الاجتماعي». واقرأ المقالة «البيانات الضخمة والتحليلات والانتخابات». وأجب عن الأسئلة التالية:

١- ما هي أنواع البيانات الضخمة التي تمَّ تحليلها في المقالة Coo؟ علِّق على بعض مصادر البيانات الضخمة.

٢- اشرح مصطلح النظام المتكامل. وما هو المصطلح الفني الآخر الذي يتناسب مع النظام المتكامل؟

٣- ما هي أنواع أساليب تحليل البيانات المُستخدَمة في المشروع؟ علِّق على بعض المبادرات التي نتجت عن تحليل البيانات.

٤- ما هي مشكلات التنبؤ المختلفة التي حلَّتها النماذج؟

٥- اذكر بعض القرارات القابلة للتطبيق التي تمَّ اتخاذها والتي تستندُ إلى نتائج التنبؤ.

٦- حَدِّد تطبيقين لتحليلات البيانات الضخمة غير المُدرَجة في المقالة.

٧- ابحث في الإنترنت عن المواد المتعلِّقة بعمل المديرين ويلعب دور تحليلات. ما هي أنواع المصادر للشركات الاستشارية والأقسام الأكاديمية والبرامج التي تجدها؟ ما هي المجالات الرئيسة التي تمَّ إعادة تصميمها؟ حَدِّد خمسة مواقع تغطي مجالاً واحداً. قدِّم نتائجك.

٨- استكشف المناطق العامة من dssresources.com. قم بإعداد قائمة بالموارد المتاحة الرئيسة. قد ترغب في الرجوع إلى هذا الموقع في أثناء عملك من خلال الكتاب.

٩- اذهب إلى microstrategy.com. قم بإيجاد معلومات عن الأنماط الخمسة من ذكاء الأعمال، قم بإعداد جدولٍ ملخصٍ لكلِّ نمط.

١٠- انتقل إلى oracle.com، وانقر على رابط Hyperion ضمن التطبيقات. حَدِّد ما هي منتجات الشركة الرئيسة. اربطها بتقنيات الدعم المذكورة في هذا الفصل.

١١- انتقل إلى موقع أسـئلة شـبكة جامعة تيراداتا. ابحث عن مقاطع فيديو BSI. راجع مقطع الفيديـو الخاص بـ «Case of Retail Tweeters». ثم قم بإعداد ملخصٍ مكوَّن من صفحة واحدة للمشـكلة والحل المقترَح والنتائج التي تمَّ تقديمها. ويمكنك أيضاً العثور على الشرائح المقترنة على slideshare.net.

١٢- قم بمراجعة قسـم النظام البيئي لبرنامج التحليـلات. وحَدِّد ما لا يقل عن شركتين إضافيتين فيما لا يقل عن خمس مجموعات من الصناعات المذكورة في المناقشة.

١٣- شملت المناقشةُ المتعلقة بالنظام البيئي التحليلي أيضاً العديدَ من عناوين الوظائف النموذجية لخريجـي برامج تحليل وعلم البيانات. ابحث عن مواقع الويب مثل datasciencecentral.com وtdwi.org لتحديـد مـا لا يقل عـن ثلاثة عناوين وظيفية مشـابهة أخرى قد تجدها مشوقةً في حياتك المهنية.

المراجع:

- Capcredit.com. (2015). "How Much Do Americans Spend on Sports Each Year?" capcredit.com/how-much-americansspend-on-sports-each-year/(accessed July 2016).
- CDC.gov. (2015, September 21). "Important Facts about Falls." cdc.gov/homeandrecreationalsafety/falls/adultfalls.html (accessed July 2016).
- CenterPointEnergy.com. "Company Overview." centerpoint energy.com/en-us/Corp/Pages/Company-overview.aspx (accessed June 2016).
- Chiguluri, V., Guthikonda, K., Slabaugh, S., Havens, E., Peña, J., & Cordier, T. (2015, June). Relationship between diabetes complications and health related quality of life among an elderly population in the United States. Poster presentation at the American Diabetes Association 75th Annual Scientific Sessions. Boston, MA.
- Cordier, T., Slabaugh, L., Haugh, G., Gopal, V., Cusano, D., Andrews, G., & Renda, A. (2015, September). Quality of life changes with progressing congestive heart failure. Poster presentation at the 19th Annual Scientific Meeting of the Heart Failure Society of America, Washington, DC.
- Davenport, T., & SAS Institute Inc. (2014, February). Analytics in sports: The new science of winning. sas.com/content/dam/SAS/en_us/doc/whitepaper2/iia-analytics-insports-106993.pdf (accessed July 2016).
- Davenport, T. H., & Patil, D. J. (2012). Data scientist. Harvard Business Review, 90, 70-76.
- Dundas.com. "How Siemens Drastically Reduced Cost With Managed BI Applications." dundas.com/resource/get case study? case Study Name = 09-03-2016-Siemens%2FDundas-BI-Siemens-Case-Study.pdf (accessed July 2016).
- Emc.com. (n.d.). "Data science revealed: A data-driven glimpse into the burgeoning new field". emc.com/collateral/about/news/emc-data-science-study-wp.pdf (accessed July 2016).
- Gartner, Inc. (2004). Using business intelligence to gain a competitive edge. A special report.
- Gates, S., Smith, L. A., Fisher, J. D., et al. (2008). Systematic review of accuracy of screening instruments for predicting fall risk among independently living older adults. Journal of Rehabilitation Research and Development, 45(8), 1105-1116.
- Gill, T. M., Murphy, T. E., Gahbauer, E. A., et al. (2013). Association of injurious falls with disability outcomes and nursing home admissions in community living older persons. American Journal of Epidemiology, 178(3), 418-425.
- Gorry, G. A., & Scott-Morton, M. S. (1971). A framework for management information systems. Sloan Management Review, 13(1), 55-70.
- Keen, P. G. W., & M. S. Scott-Morton. (1978). Decision support systems: An organizational perspective. Reading, MA: Addison-Wesley.

- Havens, E., Peña, J., Slabaugh, S., Cordier, T., Renda, A., & Gopal, V. (2015, October). Exploring the relationship between health-related quality of life and health conditions, costs, resource utilization, and quality measures. Podium presentation at the ISOQOL 22nd Annual Conference, Vancouver, Canada.
- Havens, E., Slabaugh, L., Peña, J., Haugh, G., & Gopal, V. (2015, February). Are there differences in Healthy Days based on compliance to preventive health screening measures? Poster presentation at Preventive Medicine 2015, Atlanta, GA.
- Humana. 2016 progress report. populationhealth. humana.com/wp-content/uploads/2016/05/BoldGoal2016ProgressReport_1.pdf (accessed July 2016). INFORMS. Analytics section overview. informs.org/Community/Analytics (accessed July 2016).
- NCAA, National Center for Catastrophic Sports Injury Research Report. NCAA Sport Injury fact sheets are produced by the Datalys Center for Sports Injury Research and Prevention in collaboration with the National Collegiate Athletic Association, and STOP Sports Injuries. Women's soccer injuries. ncaa.org/sites/default/files/NCAA_W_Soccer_Injuries_WEB.pdf (accessed July 2016).
- Pajouh Foad, M., Xing, D., Hariharan, S., Zhou, Y., Balasundaram, B., Liu, T., & Sharda, R. (2013). Availableto-promise in practice: An application of analytics in the specialty steel bar products industry. Interfaces, 43(6), 503-517. dx.doi.org/10.1287/inte.2013.0693 (accessed July 2016).
- Price Waterhouse Coopers Report. (2011, December). Changing the game: Outlook for the global sports market to 2015. pwc.com/gx/en/hospitality-leisure/pdf/changing-the-game-outlook-for-the-global-sportsmarket-to-2015.pdf (accessed July 2016).
- Sap.com. (2013, October). A "Smart" Approach to big Data in the Energy Industry. sap.com/bin/sapcom/cs_cz/downloadasset.2013-10-oct-09-20.a-smart-approach-to-big-data-in-the-energy-industry-pdf.html (accessed June 2016).
- Sharda, R., Asamoah, D., & Ponna, N. (2013). Research and pedagogy in business analytics: Opportunities and illustrative examples. Journal of Computing and Information Technology, 21(3), 171-182.
- Siemens.com. About Siemens. siemens.com/about/en/(accessed July 2016).
- Silvaris.com. Silvaris overview. silvaris.com/About/(accessed July 2016).
- Simon, H. (1977). The New Science of Management Decision. Englewood Cliffs, NJ: Prentice Hall.
- Tableau.com. Silvaris augments proprietary technology platform with Tableau's real-time reporting capabilities. tableau.com/sites/default/files/case-studies/silvaris-businessdashboards_0.pdf (accessed July 2016).
- teradatauniversitynetwork.com. (2015, Fall). BSI: Sports analytics—Precision football (video). teradatauniversity network.com/About-Us/Whats-New/BSI-Sports-Analytics-Precision-Football/(accessed July 2016).

الفصل الثاني
التحليلات الوصفية (١)
طبيعة البيانات، النمذجة الإحصائية، تصوير البيانات

أهداف التعلُّم:

- فهم طبيعة البيانات من حيث صِلتها بذكاء الأعمال والتحليلات.
- تعلُّم الطرق المُستخدَمة لتجهيز تحليلات البيانات الواقعية.
- وَصْف النمذجة الإحصائية وعلاقتها بتحليلات الأعمال.
- تعلُّم الإحصاء الوصفي والاستدلالي.
- تعريف تقارير الأعمال، وفهم تطوُّرها التاريخي.
- فهم أهمية تصوير البيانات/ المعلومات.
- تعلُّم أنواع مختلفة من تقنيات التصوير البياني.
- تقدير القيمة التي تضيفها التحليلاتُ المرئية إلى تحليلات الأعمال.
- معرفة قدرات وقيود لوحات المعلومات.

في عــصر البيانات الضخمة وتحليلات الأعمال التي نعيــشُ فيها؛ لا يمكن إنكار أهمية البيانات. وهناك عبارات مُســتَحدَثة مثل: «البيانات هــي النفط»، «البيانات هي اللحم الجديد»، «البيانات هي العملة الجديدة»، و«البيانات هي الملك»، هذه العبارات من شأنها أن تزيدَ من أهمية البيانات مــن جديــد. ولكن عن أيِّ نوعٍ من البيانات نتحدث؟ من الواضــح أننا لا نتحدث عن أيِّ بيانات. إنَّ مفهوم/ مبدأ المدخلات السيئة ينتجُ عنه مخرجات سيئة "garbage in garbage out-GIGO" ينطبق على ظاهرة «البيانات الضخمة» في الوقت الحالي أكثر من أيِّ بيانات في أيِّ وقتٍ سابق. ولكي تتمكَّن البياناتُ من الوفاء بوعدها وقيمة اقتراحها وقدرتها على التحوُّل إلى رؤية واقعية؛ فلا بد من الاعتناء بإنشــائها/ تحديدها، وكذلك جمعها وتكاملها وتنظيفها وتحويلها ووضعها بصورةٍ صحيحة في السياق المناسب لاستخدامها في اتخاذ القرارات بدقة وفي الوقت المناسب.

تُعَدُّ البياناتُ الموضوعَ الرئيس لهذا الفصل. ووفقًا لذلك؛ يبدأ الفصل بوَصْفٍ لطبيعة البيانات: ماهيــة البيانــات؟ وما هي الأنواع والصــور المختلفة التي يمكن أن تظهر بهــا البيانات؟ وكيف

يمكـن معالجتها مسـبقًا وتجهيزها للتحليـلات؟ يتمُّ تخصيصُ الأجـزاء القليلة الأولى من الفصل لفهم ومعالجة البيانات بشـكلٍ عميقٍ وضروري. كما تقوم أجزاءٌ قليلة بعد ذلك بوَصْف الطرق الإحصائية المُستَخدَمة لتحضير البيانات كمدخلات لإنتاج كلٍّ من المقاييس الوصفية والاستدلالية. ثـم يلي الأجزاء الإحصائية أجـزاءٌ تتعلَّق بالتقارير وتصويـر البيانات. يُعَدُّ التقريـرُ أداةَ اتصالٍ تـمَّ إعدادُها لغرضٍ مُحدَّد، وهو تحويل البيانـات إلى معلومات ومعرفة وترحيل تلك المعلومات بتنسـيقٍ قابلٍ للفهم/ الاسـتيعاب. في الوقت الحـالي، يتمُّ الاهتمام أكثر بتوجيـه هذه التقارير بصـورة مرئية، وغالبًا ما يتمُّ ذلك باسـتخدام الألوان والرسـوم البيانية، والتي تبدو بشـكلٍ عام كلوحة معلومات الغرض منها تحسينُ محتوى تلك المعلومات. وبالتالي؛ تمَّ تخصيصُ الجزء الأخير من الفصل للأقسـام الفرعية التي تُعنَى بعرض التصميم والتنفيذ، وكذلك الطريقة المثلى لتصوير المعلومات والشرح التفصيلي وعرض لوحات تلك المعلومات.

٢-١ مقـال افتتاحـي: قيـام الـ SiriusXM بجذب وإشـراك جيـلٍ جديدٍ من مستخدمي الراديو باستخدام التسويق المبني على البيانات:

تُعَـدُّ المحطـةُ الإذاعية الفضائيـة SiriusXM أكبرَ شركةٍ إذاعية في العـالم؛ إذ تُحقق عائدات سـنوية تبلغ ٣,٨ مليار دولار، وتقوم ببَثِّ مجموعةٍ واسـعةٍ من الموسـيقى الشـعبية والرياضة والأخبار والحديث وفقرات الترفيه. وقد بدأت البثَّ في عام ٢٠٠١ بعددٍ ٥٠٠٠٠ مشـترك، إلى أن وصلت إلى ١٨,٨ مليون مشترك في عام ٢٠٠٩، واليوم لديها ما يقربُ من ٢٩ مليون مشترك.

يرجع الجزءُ الأكبرُ من النموِّ الذي حققته شركة SiriusXM حتى الآن إلى الترتيبات الإبداعية مع شركات تصنيع السـيارات. واليوم تمَّ تمكينُ ما يقارب من ٧٠ ٪ من السـيارات الجديدة من شركـة SiriusXM، ومـع ذلك؛ فقد امتـدَّ أثرُ الشركة لما هو أبعد من أجهـزة الراديو في الولايات المتحدة ألا وهو التواجدُ العالمي على الإنترنت والهواتف الذكية وأيضًا التواجد من خلال خدمات وقنوات توزيع أخرى، بما في ذلك SONOS، وJetBlue، وDish.

تحَدِّي الأعمال:

عـلى الرَّغم من هـذه النجاحات الرائعة؛ فإنه في غضون السـنوات القليلـة الماضية ونتيجةً للتغـيُّرات التي طرأت على كلٍّ من التركيبة السـكانية للعملاء، والتقنية، وأيضًا حصة المنافسـين من السـوق؛ فقد أدَّى ذلك كله إلى ظهور سلسلةٍ جديدةٍ من تحَدِّيات الأعمال والفرص التجارية لـشركة SiriusXM، ونعرض فيما يلي بعضاً من أبرز تلك التحديات والفرص:

- مع زيادة مُعدَّل انتشار السيارات الجديدة في السوق؛ فقد تغيَّرت التركيبةُ السكانية للمشترين؛ فقد أصبحت أصغر سـنًا، مع دخل تقديري أقل. كيف يمكن هنا أن تتعرف شركة SiriusXM على هذه التركيبة السكانية الجديدة؟
- مع استعمال السياراتِ الجديدة ومن ثمَّ بيعها كسيارات مستعملة لمالكين جُدد، كيف يمكن هنا لـ SiriusXM أن تحَدِّد هؤلاء المالكين الجدد، وتحوِّلهم إلى عملاء مباشرين؟
- تمكَّنَت شركة SiriusXM من تقديم خدماتها عبر كلٍّ من الأقمار الصناعية والشبكات اللاسلكية؛ وذلـك بعد اسـتحواذها على أعمال المركبات المترابطة من شركـة Agero - والتي تُعتبَر الشركة الرائدة في مجال تقنيات المعلومات في سـوق السـيارات بالولايات المتحدة الأمريكية - فكيف يمكنها استغلالُ هذه الصفقة بنجاح للحصول على مصادرجديدة للدخل؟

الحلُّ المُقترَح: تحويل الرؤية نحو التسويق المستند إلى البيانات:

اعترفت شركة SiriusXM أنه لمواجهة هذه التحَدِّيات؛ فإنَّ عليها أن تصبح منظمة تسـويق عالية الأداء ومستندةً على البيانات. بدأت الشركة في إجراء هذا التحوُّل من خلال إنشاء ثلاثة مبادئ أساسية:

- أولًا: التفاعلات الشـخصية - وليس التسويق الجماعي - من شأنها أن تحكم اليوم. وسرعان ما فهمت الشركةُ أنه لإجراء المزيدِ من التسويق الشخصي؛ فعليها الاعتماد على التاريخ والتفاعلات السابقة، وكذلك على الفهم الشديد لموقع المستهلك من دورة حياة الاشتراك.
- ثانيًا: للوصول إلى هذا القدر من الفهم؛ ستحتاج تقنية المعلومات وشركاؤها في التقنية الخارجية إلى القدرة على تقديم بيانات متكاملة، وتحليلات متقدِّمة، ومِنصات تسويق متكاملة، وأنظمة توصيل مُتعدِّدة القنوات.
- ثالثًا: قد لا تتمَكَّن الشركة من تحقيق أهدافها التجارية دون وجود وجهة نظر متكاملة ومتسـقة بها، والأهم من ذلك؛ يجب أن تتحول الجوانب التقنية والتجارية في SiriusXM إلى شركاء حقيقيين للتصَدِّي على نحو أفضل للتحديات التي تواجهها في أن تصبح منظمةً تسويقيةً عالية الأداء ومعتمدةً على الأفكار المستندة على البيانات للتحدُّث مباشرةً مع المستهلكين بطرقٍ ملائمة بشكلٍ ملحوظ.

تلك الأفكارُ المبنية على البيانات، على سبيل المثال، ستُمكِّن الشركة من التفريق بين المستهلكين والمُلاك والسائقين والمستمعين وأصحاب الحسابات. كما ستساعدُ هذه الأفكار شركة SiriusXM على فهم تلك المركبات والخدمات الأخرى التي تُشكِّل جزءًا من كلِّ أسرة، وأيضًا صُنع فرص جديدة للمشـاركة. بالإضافة إلى ذلك، ومن خلال إنشـاء رؤية شاملة وموثوقة بزاوية ٣٦٠ درجة لجميع

مستهلكيها، تستطيع SiriusXM ضمان أن تكون جميع الرسائل في جميع الحملات والتفاعلات مُصمَّمةً وملائمةً ومتسقةً عبر جميع القنوات. والنقطة المهمة أيضًا هي أن يكون التسويقُ الأفضل تصميمًا وفعاليةً هو عادةً الأقل تكلفة.

التنفيذ: إنشاء ومتابعة المسار المؤدي إلى التسويق عالي الأداء:

في الوقت الذي قرّرت فيه أن تصبح شركةُ تسويقٍ عاليةَ الأداء؛ كانت SiriusXM تعمل مع منصة تسويق خارجية لا تمتلك القدرة على دعم طموحات SiriusXM؛ فاتخذت الشركة قرارًا مستقبليًا مهمًا يهدف إلى رفع قدراتها التسويقية الداخلية، ومن ثم رسمت بعناية ما يلزم القيامُ به لتحقيق ذلك بنجاح.

١- تحسين نظافة البيانات من خلال تحسين إدارة البيانات الرئيسية والحوكمة. وعلى الرغم من نفادِ صبر الشركة في وَضْع الأفكار موضع التنفيذ؛ فإنَّ نظافة البيانات كانت خطوةً أولى وضروريةً لإنشاء نافذة موثوقة في سلوك المستهلكين.

٢- نقل نشاط التحليلات التسويقية إلى داخل الشركة وتوسيع حجم البيانات؛ لرفع كفاءة تلك التحليلات وتقديم دعمٍ كاملٍ ومتكاملٍ لها.

٣- تطوير نماذج جديدة للتصنيف والتسجيل لتفعيل قاعدة البيانات، والقضاء على التأخر في الاستجابة وتكرار البيانات.

٤- توسيع حجم البيانات المتكاملة لتشمل بيانات التسويق والتسجيل، والاستفادة من تحليلات قاعدة البيانات.

٥- اعتماد مِنصة تسويقية لتطوير الحملات.

٦- وَضْع كلِّ تلك القدرات معًا؛ لتحقيق إدارة العروض في وقتٍ ملائمٍ عبر جميع قنوات التسويق: مركز الاتصال، والتليفونات الخلوية، ومواقع الإنترنت، والتطبيقات الجاهزة.

إنَّ إكمال كلِّ تلك الخطوات يعني العثور على شريك التقنية المناسب. فوقع اختيار SiriusXM على تيراداتا؛ إذ شكَّلت نقاطُ قوتها رابطةً قويةً بين المشروع والشركة؛ فكانت تيراداتا قادرةً على القيام بما يلي:

- دَمْج مصادر البيانات مع مستودعِ بيانات متكامل (IDW)، وتحليلات متقدمة، وتطبيقات تسويق قوية.
- حلُّ المشكلات وقت استجابة البيانات.
- تقليص معنوي في حركة البيانات عبر قواعد بيانات وتطبيقات مُتعدِّدة.
- سلاسة التفاعل مع التطبيقات الجاهزة والوحدات النمطية لجميع مجالات التسويق.

- القياس والتنفيذ بمستويات عالية جدًا؛ بهدف تشغيلِ الحملات والتحليلات ضِمن قاعدة البيانات.
- فتح قنوات اتصالٍ في وقتٍ ملائم مع العملاء.
- تقديم الدَّعم التشغيلي (سواءً عبر السَّحابة أو في أماكن العمل).

وقد مكَّنَت شركة تيراداتا شركة SiriusXM من التحرُّك بسلاسة وبسرعة على امتداد خارطة الطريق التي رسمتها لنفسها، والشركة الآن في خضم عملية تحوُّلية مدتها خمس سنوات. وبعد تأسيسها لعملية إدارة البيانات القوية، بدأت SiriusXM من خلال تنفيذ مستودع بيانات متكاملة لها، والذي مكَّن الشركة بسرعة وبطريقة فعَّالة في تفعيل رؤيتها الجديدة في جميع أنحاء المنظمة.

قامت الشركة بعد ذلك بتنفيذ ما يُسمَّى "مدير التفاعل مع العملاء" - وهو جزءٌ من برنامج تيراداتا Integrated Marketing Cloud - الذي يتيحُ التفاعل حواريًّا مع العميل في الوقت الملائم؛ وذلك عبر مجموعةٍ كاملةٍ من قنوات الاتصال الرقمية والتقليدية، وأيضًا يسمح بالتعاون بين SiriusXM ومركز الرسائل الرقمية لـ تيراداتا.

وستتيح تلك الحزمة من الإمكانيات المجتمعة لشركة SiriusXM التعامُلَ مع الاتصالات المباشرة عبر قنوات مُتعددة. كما أن هذا التطوُّر سوف يتيحُ العروض في الوقت الملائم، وكذلك الرسائل التسويقية والتوصيات بناءً على السلوك السابق.

وبالإضافة إلى تبسيطها لكيفية تنفيذ وتحسين الأنشطة التسويقية الصادرة؛ تتحكَّم SiriusXM أيضًا في عمليات التسويق الداخلية الخاصة بها بالتنسيق مع إدارة الموارد التسويقية، والذي أيضًا يُعَدُّ جزءًا من تيراداتا Integrated Marketing Cloud. وسيسمح هذا الحلُّ لشركة SiriusXM بتبسيط سَيْر العمل، وتحسين الموارد التسويقية، ورفع الكفاءة من خلال كلِّ قرش من ميزانية التسويق الخاصة بهم.

النتائج: حَصْد الفوائد:

بينما الشركةُ تواصل تطوُّرها لتصبح منظمةً تسويقيةً عالية الأداء؛ إذ SiriusXM تستفيدُ بالفعل من إستراتيجيتها المنفَّذة بعناية. تعملُ كلٌّ من إحصاءات المستهلك على مستوى الأسرة والرؤية الكاملة لإستراتيجية اللمس التسويقية مع كلِّ مستهلكٍ على تمكين SiriusXM من إنشاء عروضٍ أكثر استهدافًا على مستوى كلٍّ من الأسرة والمستهلكين والأجهزة. ومن خلال جعل قدرات البيانات والتحليلات التسويقية من ضِمن أنشطة الشركة الداخلية، تمكَّنت SiriusXM من تحقيق الآتي:

- الحصول على نتائج الحملة فورًا تقريبًا بدلًا من ٤ أيام؛ مما يؤدي إلى تخفيضات هائلة في مُعدَّل الدوران الزمني للحملات والمحلِّلين الداعمين لها.

- الحصــول على رؤية دائرية مُغلقة تســمح للمحلِّلين بدعم الحــوارات المتعددة المراحل وإجراء تعديلات داخل الحملة لزيادة فعاليتها.
- الحصول على نمذجةٍ وتســجيلٍ فوري؛ لرفع مســتوى الذكاء التسويقي وشَحْذ عروض الحملات والاستجابات بسرعة تلائم أنشطة الشركة التجارية.

وأخيرًا، عزَّزَت تجربة SiriusXM فكرةَ أنَّ التسويق العالي الأداء يُعَدُّ مفهوماً يتطور باستمرار. كما قامت الشركة بتنفيذ كلٍّ من العمليات والتقنية التي تمنحها القدرة على النموِّ المستمرِّ والمرن.

أسئلة على المقالة الافتتاحية:

١- ماذا تفعل SiriusXM؟ وفي أيِّ نوعٍ من السوق تقوم بأعمالها؟

٢- ما التحَدِّيات التي واجهتها؟ مع التعليق على كلٍّ من التقنية والتحَدِّيات المتعلقة بالبيانات.

٣- ما الحلول المقترحة؟

٤- كيف قامت الشركةُ بتنفيذ الحلول المقترحة؟ وهل واجهوا أيَّ تحديات في أثناء التنفيذ؟

٥- ماذا كانت النتائج والفوائد؟ وهل كانت تستحقُّ الجهد/ الاستثمار؟

٦- هــل يمكنك التفكــير في شركاتٍ أخرى تواجه تحَدِّياتٍ مماثلة، ويمكنها الاســتفادة من الحلول التسويقية المشابهة القائمة على البيانات؟

ماذا يمكننا أن نتعلَّم من هذه المقالة القصيرة؟

في ســعيها إلى النجاح ضِمن بيئة صناعة تنافســية سريعة التغيُّر؛ أدركت SiriusXM الحاجة إلى بنية تحتية تســويقية جديدة ومُطوَّرة (والتي تستندُ إلى البيانات والتحليلات)؛ وذلك لتوصيل القيمة المقترحة بفعالية لعملائها الحاليين والمحتملين. وكما هو الحالُ في أيِّ صناعة، أو في الترفيه؛ يعتمــد النجاحُ أو البقاء طويلًا على استشــعار النزعات المتغيرة (التفضيــلات وعدم التفضيلات) بذكاء عالٍ، ووَضْع الرَّســائل الصحيحة والسياســات المناســبة معًا؛ بهدف كسب عملاء جُدد مع الإبقاء على العملاء الحاليين. الغرض هنا؛ هو إنشــاء وإدارة الحملات التســويقية الناجحة التي يتردد صداها مع الســكان المُستَهدَفين من العملاء الذين يعكسون حُزمة من ردود الأفعال وثيقة الصِّلــة؛ تهــدف إلى ضبط وتعديل الرســالة من أجل تحســين المخرجات. في النهايــة؛ كان الأمرُ يتعلــق كليًّا بدقة الطريقة التي تمَّ اتباعها في الأنشــطة التجارية، وهي أن تكون ســبَّاقًا بشــأن الطبيعة المتغيرة للعملاء، وأيضًا توفير المنتجات والخدمات المطلوبة في الوقت المناســب للعميل؛

وذلك باســتخدامٍ كاملٍ لإستراتيجية التســويق القائمة/ المعتمدة على البيانات. وقد لعب كلٌّ من تحديدِ وإنشاء المصادر والوصول إليها وجمعها، وتكاملها وتنظيفها وتحويلها وتخزينها ومعالجتها للبيانات محلَّ الدراســة دورًا مهــمًّا في نجاح SiriusXM في تصميم وتنفيذ إســتراتيجية تحليلية تســويقية، كما هو الحال في أيِّ شركة ناجحة في الوقت الحاضر باســتخدام الذكاء التحليلي، بغض النظر عن الصناعة التي يشاركون فيها.

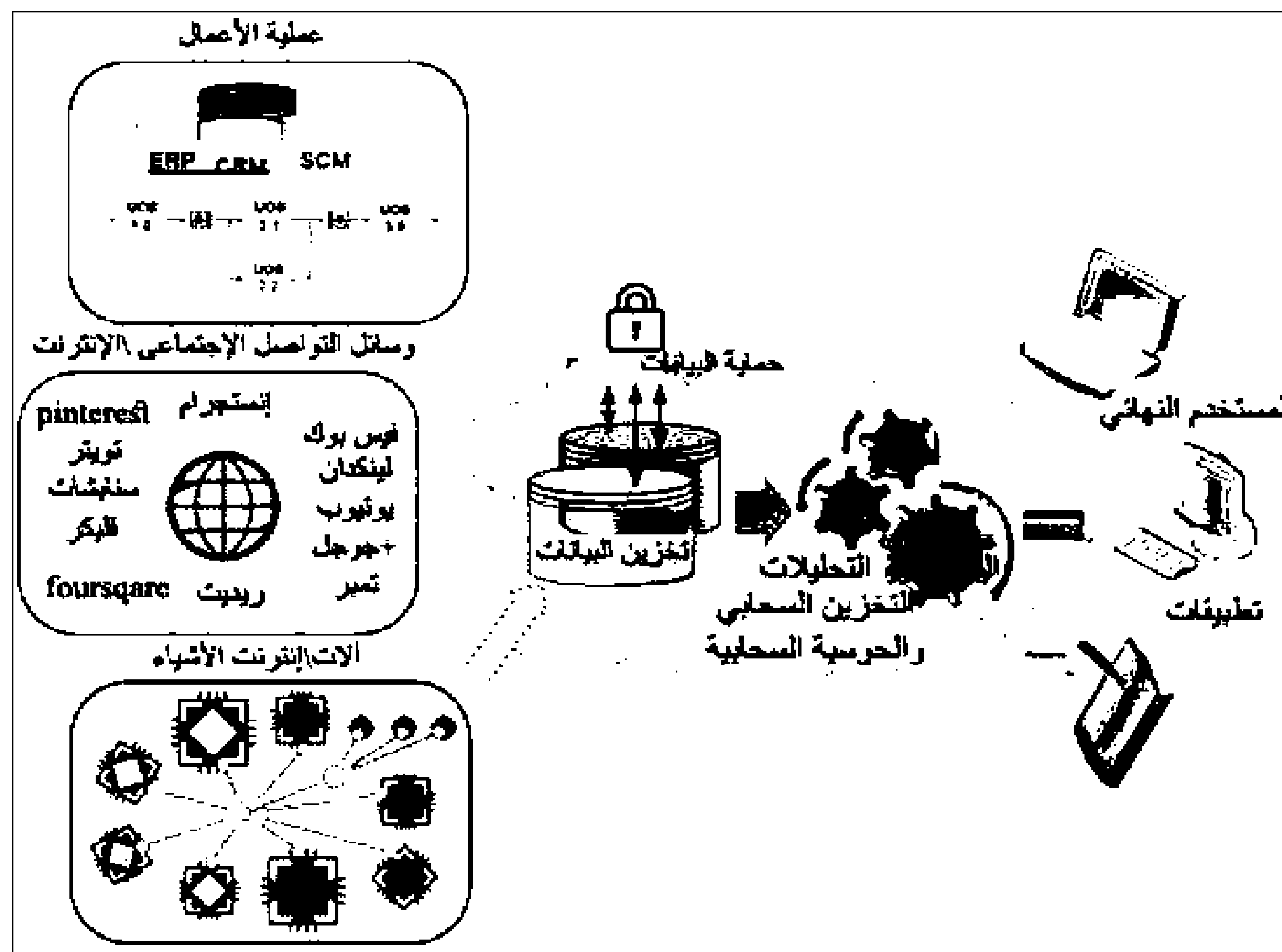
٢-٢ طبيعة البيانات:

تُعَدُّ البيانات هي العنصر الرئيسي لأيِّ مبادرة (ذكاء أعمال)، وعلم البيانات، وتحليلات الأعمال. في الواقــع؛ يمكِن اعتبارُ البيانات أنها المادة الخام لما تنتجه هذه التقنيات الشــائعة لاتخاذ القرار مــن معلومــات ورؤية ومعرفة. وبدون بيانات ما كان لأيٍّ من هذه التقنيات أن تظهر أو تنتشر، وعلى الرغم من أننا قمنا - وبشكلٍ تقليدي - ببناء نماذج تحليلية باستخدام كلٍّ من معرفة وخبرة الخبراء ببيانات قليلة جدًا أو بدون بيانات على الإطلاق، على كلِّ حالٍ كان ذلك في السابق أمَّا الآن فقد أصبحت البيانات ركيزةً أساسيةً في كلِّ ذلك. ونظراً لكون جَمْع وتخزين وإدارة البيانات من أكبر التحَدِّيات؛ فإنها تُعَدُّ (أيْ: البيانات) ضِمن أغلى وأثمن أصول أيِّ منظمة، مع إمكانية إنشــاء رؤية لا تُقدَّر بثمن لفهم العملاء والمنافسين وعمليات الأعمال بشكلٍ أفضل.

يمكِــن أن تكون البياناتُ صغيرةً ويمكِن أن تكون كبيرةً جــدًا، ومن الممكِن أن تكون البيانات مُهيكلةً (مُعدَّة بصورة جيدة للتعامل مع أجهزة الحاســب)، أو يمكِن أن تكون غير مهيكلة (على سبيل المثال: يتمُّ إنشاءُ نصٍّ يدويٍّ، وبالتالي لا يمكن فهمه/ التعامل معه بواسطة أجهزة الحاسب). ومــن الممكِن أن تــأتي البيانات دائمًا في مجموعاتٍ صغيرة أو تتقولــب دفعة واحدة في مجموعة أكبر. هناك بعضُ الخصائص التي تُحدِّد الطبيعة المتأصلة للبيانات الآن، والتي نسميها في الغالب البيانــات الضخمــة أو الكبيرة الحجم. وعلى الرغم من أن هــذه الخصائص للبيانات تجعلها أكثرَ صعوبةً في المعالجة والاســتخدام؛ فإنهــا تزيدُ من قيمتها؛ لأنها تجعلها أكثر إثراءً للدراســة بعيدًا عن مفاهيمها المألوفة؛ مما يســمح باكتشــاف معرفةٍ جديدةٍ ومبتكرة. لقد انقضى زمنُ الطرق

Sources: Quinn, C. (2016). Data-driven marketing at SiriusXM. Teradata Articles & News. at http://bigdata.teradata.com/US/Articles-News/Data-Driven-Marketing-At-SiriusXM/(accessed August 2016); Teradata customer success story. SiriusXM attracts and engages a new generation of radio consumers. http://assets.teradata.com/resourceCenter/downloads/CaseStudies/EB8597.pdf?processed=1.

التقليديـــة لجمع البيانات يدويًّا (إمَّا عن طريق الاســتطلاعات، أو عبر معاملات تجارية مباشرة) وحلَّ محلَّها آلياتٌ جديدةٌ لجمع البيانات الحديثة التي تستخدم الإنترنت و/أو أجهزة الاستشعار/ RFID المعتمدة على الشــبكات المحوسبة. إن هذه الأنظمة الآلية لجمع البيانات لا تمكِّننا فقط من جمع مزيدٍ من البيانات؛ ولكن أيضًا ترفع من جودة البيانات، وتؤكِّد على ســلامتها. ويُوضِّح الشكل ٢-١ سلسلة التحليل النموذجية - بيانات إلى تحليلات إلى معلومات قابلة للتنفيذ.



شكل ٢-١: بيانات الاستمرارية والمعرفة

وعــلى الرَّغم مــن المكانة الثمينة التي تحتلها البيانات والتــي لا يمكن إنكارُها؛ فإنه يجبُ أن تلتزمَ البياناتُ ببعض القياســات الأساســية ذات الجودة والقابلية للاســتخدام. ومن المعلوم أنه ليســت كلُّ البيانات مفيدةً لجميع المهام؛ إذ يجبُ أن تتطابق البياناتُ مع المهمة المراد استخدام تلــك البيانات فيها بجانب تغطية التفاصيل الخاصة بتلك المهمة. وبالنســبة لأيِّ مهمة مُحدَّدة؛ يجبُ أن تتوافق البيانات المتاحة مع متطلبات الجودة والكمية. وبشكلٍ أساسي؛ لا بد من تجهيز البيانات للتحليلات. والسؤال هنا ما هو تجهيز البيانات للتحليلات؟ فبالإضافة إلى ضرورةِ ملاءمتها للمشــكلة محلَّ البحث ومتطلبات الجودة/الكمية؛ يجب أن يكون للبيانات هيكلٌ معينٌ يتماشى مع الحقول/المتغيرات الرئيســة مع القيم الطبيعية الصحيحــة. علاوةً على ذلك؛ يجب أن يكون هناك تعريفٌ متفقٌ عليه على مســتوى المؤسسة للمتغيرات والموضوعات الشائعة (أحيانًا تُسمَّى

أيضًا إدارة البيانات الرئيسـة)، مثل كيفية تعريف العميل (ما هي خصائص العملاء المُستَخدَمة لإنتاج تمثيلٍ شاملٍ بما يكفي للتحليلات) وفي أيٍّ مرحلة من النشاط التجاري يتمُّ الحصولُ على المعلومات المتعلقة بالعميل والتحقُّق منها وتخزينها وتحديثها.

في بعض الأحيان؛ قد يعتمدُ تمثيلُ البيانات على نوع التحليلات المُستَخدَمة. عادةً ما تتطلَّب الخوارزميات التنبؤية بشكلٍ عام وجودَ ملفٍ ثابتٍ مع متغيرٍ مُستَهدَف؛ لذا فإنَّ تجهيز تحليلات البيانات للتنبؤ؛ يعني ضرورةَ تنسيق مجموعات البيانات على هيئة ملف مُسطَّح مع تجهيزها للدخول في تلك الخوارزميات التنبؤية. كما أنه من الضروري أيضًا مطابقة البيانات مع متطلبات برمجيات خوارزمية التنبؤ - على سبيل المثال: تتطلَّب خوارزميات الشبكة العصبية التمثيل العددي لكلِّ المتغيرات المُدخلَة (حتى إنَّ المتغيرات الاسمية تحتاج إلى تحويلها إلى متغيرات ثنائية رقمية زائفة؛ في حين لا تتطلَّب خوارزميات شجرة القرار مثل هذا التحويل الرقمي)، والتعامل بسهولةٍ مع مزيجٍ من المتغيرات الاسمية والرقمية.

غالبًا ما تَصِل مشروعاتُ التحليلات التي تتجاهل المهامَّ المتعلقة بالبيانات إلى إجاباتٍ غير ملائمة للمشكلة محل البحث، وقد تؤدي هذه الإجابات التي تمَّ الوصولُ إليها عن غير قصدٍ، والتي قد تبدو جيدةً، إلى اتخاذ قرارات غير دقيقة وغير ملائمة. ونعرضُ فيما يلي بعض الخصائص (المقاييس) التي تُحدِّد مستوى جاهزية البيانات لإجراء دراسة تحليلية (Delen، 2015، Kock وMcQueen وCorner، 1997).

- **موثوقية مصدر البيانات:** يشيرُ إلى أصالة وملاءمة وسيط التخزين الذي يتمُّ الحصولُ على البيانات من خلاله - والسؤال هنا هو "هل لدينا ثقةٌ تامةٌ في مصدر البيانات هذا؟"، إذا كان ذلك ممكنًا؛ فيجبُ علينا دائمًا أن نبحث عن المصدر الأصلي/المنشأ الأصلي لتلك البيانات؛ وذلك بهدف إزالة/ تقليل احتمال تشويه البيانات وتحويل البيانات الناتجة عن سوء المعالجة في أثناء نقل البيانات من مصدرها إلى وجهتها التالية على مرحلةٍ واحدةٍ أو أكثر؛ إذ إنَّ كلَّ تحرُّك للبيانات يعطي فرصةً لإسقاط أو إعادة صياغة بنود تلك البيانات عن غير قصد؛ مما يَحِدُّ من سلامة مجموعة البيانات وربما دقتها الحقيقية.
- **دقة محتوى البيانات:** يعني أنَّ البيانات صحيحةٌ ومطابقةٌ بشكلٍ جيدٍ لمشكلة التحليلات - وهنا سؤالٌ آخر "هل لدينا البيانات الصحيحة للمهمة المطلوبة؟" يجبُ أن تمثل البياناتُ الهدف الذي تمَّ تحديده من خلال المصدر الأصلي للبيانات. فعلى سبيل المثال: يجب أن تكون معلوماتُ التواصل مع العميل المسجَّلة في قاعدة البيانات هي نفس ما قاله العميل شخصيًّا.

هذا وسيتمُّ تناول موضوع دقة البيانات بمزيدٍ من التفصيل في الفقرة التالية من هذا الفصل.

- **إمكانية الوصول إلى البيانات:** يعني أن البيانات يمكِن الحصول عليها بسهولةٍ ويُسر - وذلك بالإجابة عن السؤال التالي "هل يمكِننا الوصولُ بسهولة إلى البيانات عندما نحتاج إليها؟" قد يكون الوصولُ إلى البيانات خادعًا؛ خاصةً إذا كانت مستودعات البيانات في أكثر من موقعٍ ووسط تخزين وتحتاج إلى دمج/ تحويل عند الوصول إليها والحصول عليها. ومع تقلُّص دور أنظمة إدارة قواعد البيانات العلائقية التقليدية (أو تعايشها مع) جيلٍ جديدٍ من وسائط مستودعات البيانات، مثل: بحيرات البيانات، والبنية التحتية Hadoop، تتزايد في نفس الوقت أهمية/ حساسية الوصول للبيانات.
- **حماية وخصوصية البيانات:** يعني أنَّ البيانات متاحةٌ فقط لهؤلاء الأشخاص الذين لديهم السُّلطة والحاجة إلى الوصول إليها ومَنْع أيِّ شخصٍ آخر من ذلك. إن تزايد الرواج في الدرجات التعليمية وبرامج شهادات ضمان المعلومات؛ هو دليلٌ على الحاجة الماسَّة والمتزايدة لمقياس جودة البيانات هذا. ينبغي على أيِّ منظمة طبية تحتفظ بسجلات صحية لكلِّ مريضٍ على حِدَة؛ أن يكون لديها أنظمةٌ مُطبَّقةٌ لا تقوم فقط بحماية البيانات من وصولِ أشخاص غير مُصرَّح لهم إليها (والذي يتمُّ تفويضه بموجب قوانين اتحادية، مثل قانون قابلية التأمين الصحي وقابلية التأمين [HIPPA])؛ ولكنها تُحدِّد بدقة السماح وإمكانية الوصول إلى سجلات كلِّ مريض من قِبل الأشخاص المُصرَّح لهم بذلك (Annas, 2003).
- **ثراء البيانات:** يعني أن جميعَ عناصر البيانات المطلوبة مُدرَجةٌ في مجموعة البيانات. في الأساس؛ يعني الثراء (أو الشمولية) أن المتغيِّرات المتاحة تُصوِّر بُعدًا ثريًا بما يكفي لدراسة الموضوع الأساسي دراسةً تحليليةً دقيقةً وقيِّمةً. ويعني أيضًا أنَّ محتوى المعلومات مكتملٌ (أو شِبه مكتمل) لبناء نموذج تحليلات تنبؤية و/أو توجيهية.
- **اتساق البيانات:** يعني أنَّ البيانات يتمُّ جمعها بدقةٍ وكذلك دمجها/ تركيبها. وتمثل البيانات المتسقة معلومات الأبعاد (متغيرات الاهتمام) الواردة من مصادر متفاوتة الاحتمال ولكنها تتعلق بنفس الموضوع. فإذا لم يتم دمج/ تكامل البيانات بشكلٍ صحيح؛ فقد تُوجَد بعضُ المتغيرات الخاصة بموضوعات مختلفة في سجلٍ واحد - مع وجود مجموعتين من السِّجلات المختلفة للمرضى، على سبيل المثال: قد يحدث ذلك في أثناء دمج سجلات بيانات نتائج الفحص السكانية والمرضية.
- **توقيت البيانات:** يعني أنَّ البيانات يجب أن يتمَّ تحديثها (أو تكون حديثةً/ جديدةً كما يجب أن تكون) لنموذج التحليلات المُعطى. كما يعني أن البيانات يتمُّ تسجيلها في وقت الحدث أو الملاحظة أو بالقرب منه؛ بحيث يتمُّ مَنْع التحريف المتعلق بالوقت للبيانات (التذكر والتشفير غير الصحيحين).

ونظرًا لأن التحليلات الدقيقة تعتمد على بيانات دقيقة وفي الوقت المناسـب؛ فإن السِّـمة الأساسية للبيانات الجاهزة للتحليلات هي توقيت إنشاء عناصر البيانات والوصول إليها.

- **مسـتوى تفصيل البيانـات:** يتطلَّب ذلك أن يتـمَّ تعريفُ المتغيرات وقِيـم البيانات على أدنى مسـتوى (أو منخفض بقـدر الإمكان) من التفاصيل للاسـتخدام المطلوب للبيانـات. فإذا تمَّ تجميـعُ البيانات؛ فقد لا تحتوي على مسـتوى التفاصيل اللازمـة لخوارزمية التحليلات بغرض تعلُّم طريقة تمييز سجلات/ حالات مختلفة بعضها عن بعض. على سبيل المثال: وفي إطار طبِّي، يتحتـم تقريبُ القيم العددية للنتائج المعملية إلى الخانة العشرية المناسـبة للتفسـير الهادف لنتائج الاختبار والاستخدام السليم لتلك القِيم داخل خوارزمية التحليلات. وبالمثل، في مجموعة البيانات المتعلقة بالتوزيع السُّـكاني؛ ينبغي تعريفُ عناصر البيانات على مستوًى ملائمٍ لتحديد الفـروق في نتائج الرعاية بين مختلف المجموعات السـكانية الفرعيـة. شيءٌ واحدٌ يجب علينا تذكره هنا؛ هو أنَّ البيانات المُجمَّعة لا يمكن تقسيمها (دون الوصول إلى المصدر الأصلي)، ولكن يمكن تجميعها بسهولة من تمثيلها الدقيق.
- **صلاحيـة البيانـات:** هو المصطلحُ المُسـتَخدَم لوصف التطابق/ عدم تطابق بـين قيم البيانات الفعليـة والمتوقعة لمتغيرٍ ما. وكجزءٍ من تعريـف البيانات؛ يجب تحديدُ القِيم أو نطاق القيم المقبولة لكلِّ عنصر بيانات. على سـبيل المثال: يتضمَّن تعريفُ البيانات الصحيح المرتبط بنوع الجنسِ ثلاث قيم: ذكر، وأنثى، وغير معروف.
- **ملاءمة البيانات:** يعني أنَّ المتغيرات في مجموعة البيانات كلها ذات صِلة بالدراسـة التي تتم. ولا تُعَـدُّ العلاقة مقياسًـا ثنائيَّ التفرُّع (كون المتغـير مرتبطاً بالدراسـة أم لا)، بل يكون لديها مجموعـةٌ من الروابط من الأقل صِلـة إلى الأكثر أهميةً. واسـتنادًا إلى خوارزميات التحليلات المُسـتَخدَمة، قد يختار الدارس فقط تضمين المعلومات الأكثر صِلة (أيْ: المتغيرات) أو إذا كانت الخوارزمية قادرةً على تصنيفها؛ فقد يختارُ تضمين جميع المتغيرات ذات الصِّلة بغض النظر عن مسـتوى ارتباطها. هناك أمرٌ مهمٌّ يجب على دراسات التحليلات تجنُّبه ألا وهو تضمين بيانات غـير ملائمة تمامًا في بناء النموذج؛ فقد يؤدي ذلـك إلى تلويث المعلومات الخاصة بالخوارزمية؛ مما يؤدي إلى نتائج غير دقيقة ومُضلِّلة.

وعـلى الرغم من أنَّ هـذه المقاييس قد تكون الأكثر شـيوعًا؛ فإنَّ جـودةَ البيانات الحقيقية والاسـتعداد الممتاز للتحليلات لنطاق تطبيق معين؛ سيتطلبان مستوياتٍ مختلفةً من التركيز على هـذه الأبعاد المترية، وربما إضافة المزيد من التفاصيل المُحدَّدة لهذه المجموعة. وسـوف نتناول

في الفصـل التالي طبيعةَ البيانات بمزيـدٍ من التفاصيل من منظورٍ تصنيفي لـسرد أنواع البيانات المختلفة وتحديدها من حيث صِلتها بمشاريع التحليلات المختلفة.

أسئلة مراجعة على القسم ٢-٢:

١- كيف تصِف أهمية البيانات في التحليلات؟ وهل يُمكننا التفكيرُ في التحليلات بدون بيانات؟

٢- بالنظر إلى التعريف الجديد والواسع لتحليلات الأعمال، ما هي المُدخلات والمخرجات الرئيسية لاستمرار التحليلات؟

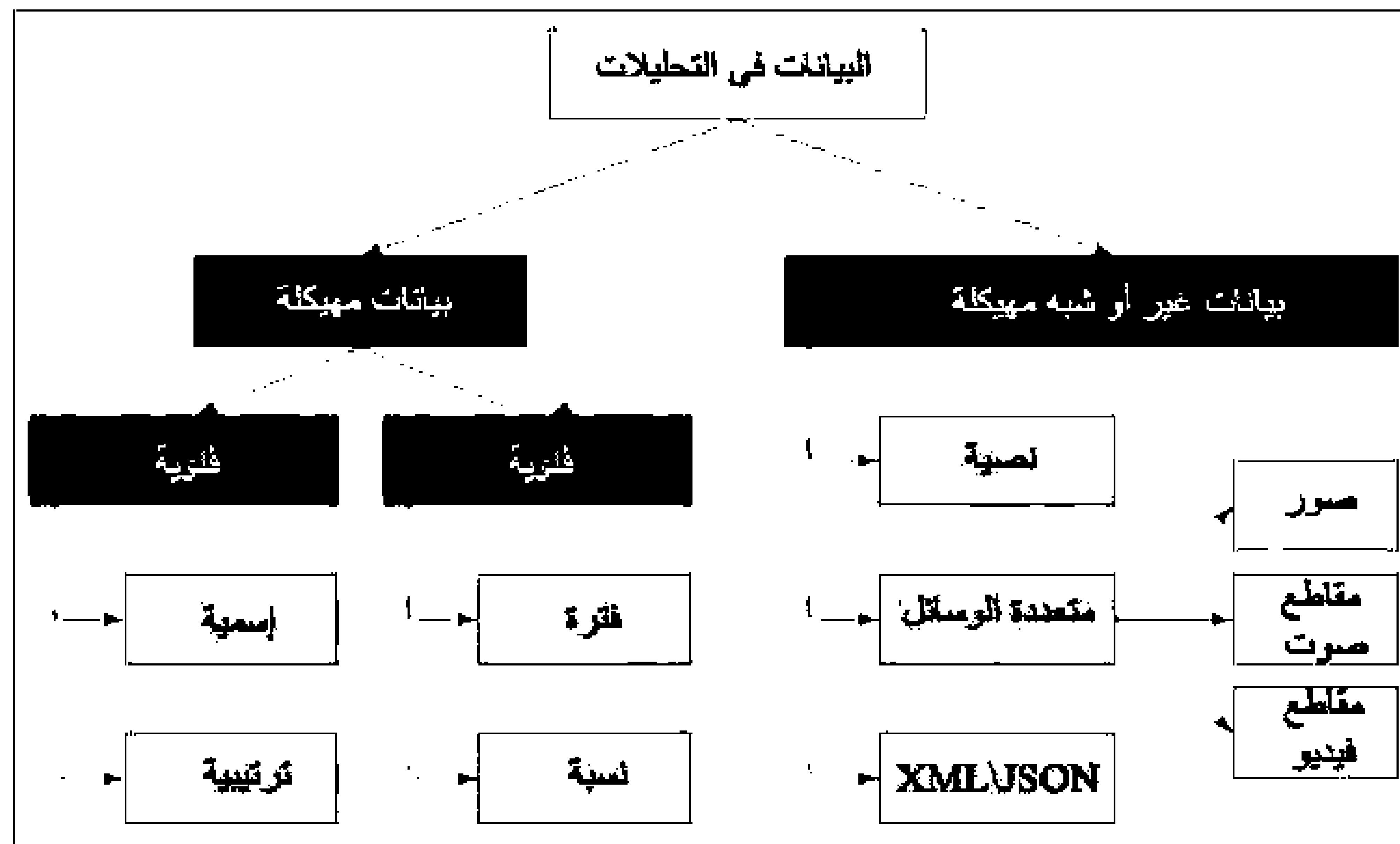
٣- من أين تأتي بيانات تحليلات النشاط التجاري؟

٤- في رأيك؛ ما أهم ثلاثة تحديات متعلِّقة بالبيانات؛ للوصول إلى تحليلاتٍ أفضل؟

٥- ما المقاييسُ الأكثر شيوعًا التي تعملُ على تجهيز البيانات للتحليلات؟

٢-٣ تصنيف مُبسَّط للبيانات:

تشـير البيانات (لا سيما في صيغتها الفردية) إلى مجموعةٍ من الحقائق التي يتمُّ الحصول عليها عـادةً كنتيجةٍ للتجارب أو المشـاهدات أو المعاملات أو الخبرات. وقـد تتكوَّن البيانات من الأرقام والحروف والكلمات والصور والتسـجيلات الصوتية وما إلى ذلك، كقياسات لمجموعة من المتغيرات (خصائص الموضوع أو الحدث محل الدراسة). فغالبًا ما يُنظر إلى البيانات على أنها أدنى مستوى من التجريد (الغموض) يتمُّ استخلاص معلومات منه، ومن ثمَّ معرفة. أمَّا في أعلى مستوى من التجريد؛ فيمكِن تصنيفُ البيانات على أنها منظمةٌ وغير منظمة (أو شبه مهيكلة). تتكون البيانات غير/ شبه المهيكلة من أيٍّ مزيجٍ من المحتوى النصِّي والصور والصوت ومحتوى الويب. وستتمُّ تغطية البيانات غير/ شبه المهيكلة بمزيدٍ من التفصيل في فصل تعدين النص وتعدين الويب. أمَّا البيانات المهيكلة؛ فهـي تلك التي تسـتخدمها خوارزميات التنقيـب في البيانات، ويمكن تصنيفها عـلى أنها فئويةٌ أو رقمية. ويمكن تقسيمُ البيانات الفئوية إلى بيانات اسمية أو ترتيبية، في حين تقسَّم البيانات الرقمية إلى فواصل زمنية أو نِسَب. يُوضِّح الشكل ٢-٢ تصنيفاً بسيطاً للبيانات.



شكل ٢-٢: تصنيف بسيط للبيانات

- **بيانات فئوية:** تمثِّل علامات الفئات المتعدِّدة المُستَخدَمة لتقسيم متغير إلى مجموعات مُحدَّدة. وتتضمَّن أمثلة المتغيرات الفئوية كالعرق والجنس والفئة العمرية والمستوى التعليمي. وبالرغم من أن المتغيرين الأخيرين قد يتمُّ النظرُ فيهما أيضًا بطريقة عددية باستخدام قِيَم دقيقة للعمر وأعلى درجة مكتملة؛ فإنه غالبًا ما يكون أكثرَ إفادةً لتصنيف مثل هذه المتغيرات إلى عددٍ صغيرٍ نسبيًّا من الطبقات المرتبة. وقد يُطلَق على البيانات الفئوية أيضًا بيانات منفصلة؛ مما يعني أنها تمثل عددًا محدودًا من القيم دون اتصال بينها. وحتى إن كانت القِيم المُستَخدَمة للمتغيرات الفئوية (أو المنفصلة) هي قِيم رقمية؛ فإنَّ هذه الأرقام ليست أكثرَ من رموزٍ ولا تشيرُ إلى إمكانية حساب قِيم كسرية.
- **بيانات اسمية:** تحتوي على قياسات الرموز البسيطة المُخصَّصة للأشياء كتسميات؛ ولكنها لا تُعتبَر قياسات. فعلى سبيل المثال: يمكِن تصنيفُ المتغير (الحالة الاجتماعية) بشكلٍ عام إلى (١) أعزب، (٢) متزوج، (٣) مطلق. كما يُمكِن تمثيلُ البيانات الاسمية ذات القيم الثنائية بقيمتين محتملتين (على سبيل المثال: نعم/ لا، أو صواب/ خطأ، أو جيد/ سيء، أو ما إلى ذلك)، أو ذات القيم المتعددة بثلاث قيم محتملة أو أكثر (مثل: بني/ أخضر/ أزرق، أو أبيض/ أسود/ لاتيني/ آسيوي، أو أعزب/ متزوج/ مطلق، أو ما إلى ذلك).
- **البيانات الترتيبية:** تحتوي على رموزٍ مُخصَّصة للأشياء أو الأحداث كتسمياتٍ، وفي نفس الوقت تُعبِّر أيضًا عن رُتبة الترتيب فيما بينهم. فعلى سبيل المثال: يمكِن تصنيفُ المتغير (درجة

الائتمان) بشكلٍ عام إلى (١) منخفضة، أو (٢) متوسطة، أو (٣) مرتفعة. كما يمكن وَضْع العلاقات المرتبة المماثلة في متغيرات، مثل: الفئة العمرية (بمعنى: طفل، شاب، متوسط العمر، مسن)، والمستوى التعليمي (بمعنى: المدرسة الثانوية، الكلية، الدراسات العليا).

- **البيانات الرقمية:** تمثل القيمُ العددية لمتغيراتٍ مُحدَّدة. وتشمل أمثلة المتغيرات ذات القيمة العددية السِّن، وعدد الأطفال، ومجموع دخل الأسرة (بالدولار الأمريكي)، ومسافة السفر (بالأميال)، ودرجة الحرارة (بمقياس فهرنهايت). كما يمكن أن تكون القِيم العددية للمتغير عددًا صحيحًا (بلا كسور) أو حقيقية (تشمل الصحيحة بالإضافة إلى الكسور). وقد تُسمَّى البياناتُ الرقمية أيضًا بيانات مستمرة؛ مما يعني أنَّ المتغير يحتوي على مقادير متواصلة على مقياس مُحدَّد بما يسمحُ بإدخال قيم مؤقتة. فبخلاف المتغير المنفصل الذي يمثل بيانات محدودة وقابلة للعد، يمثل المتغير المستمر قياسات قابلة للتطوير، ومن الممكن أن تحتوي البيانات على عددٍ لا نهائي من القِيَم الكسرية.
- **بيانات فترة:** هي مُتغيرات يُمكِن قياسها على مقاييس الفواصل الزَّمنية. هناك مثالٌ شائعٌ لقياس الفواصل الزمنية؛ هو درجة الحرارة المئوية. في هذا المقياس المُحدَّد، تكون وحدة القياس ١/ ١٠٠ من الفرق بين درجة حرارة الانصهار ودرجة حرارة الماء المغلي في الضغط الجوي؛ بما يعني عدمَ وجود قيمة صفرية مُطلَقة.
- **بيانات نسبية:** وتشمل المتغيرات الشائعة للقياس والموجودة في علوم الفيزياء والهندسة. ويُعَدُّ كلٌّ من المجموعة والطول والوقت وزاوية الطائرة والطاقة والشُّحنة الكهربائية أمثلةً على القياسات الفيزيائية التي هي مقاييس النِّسَب. ويَستمدُّ هذا النوع من المقاييس اسمَه من حقيقة أنَّ القياس هو تقديرُ النسبة بين حجم الكمية المستمرة وحجم وحدةٍ من نفس النوع. وبشكلٍ غير رسمي؛ تُعَدُّ السِّمة المميزة لمقياس النسبة هي وجودُ خانةٍ صفرية القيمة وغير عشوائية. فعلى سبيل المثال: فإنَّ مقياس درجة الحرارة كلفن له نقطة صفر غير عشوائية من الصفر المطلق، والتي تساوي (-٢٧٣,١٥) درجة مئوية. تُعَدُّ نقطة الصفر هذه غير عشوائية؛ لأن الجسيمات التي تشكِّل المادة عند هذه الدرجة لها طاقةٌ حركيةٌ صفرية.

يجب تحويلُ أنواع البيانات الأخرى، كبيانات النصوص والحيِّز والصور والفيديو والصوت، إلى شكلٍ من أشكال التمثيلِ الفئوي أو الرقمي قبل أن تتمَّ معالجتُها بواسطة طرق التحليل (خوارزميات التنقيب في البيانات؛ Delen, 2015). كما يمكِن أيضًا تصنيفُ البيانات إلى بيانات ثابتة أو حركية (أيْ: سلاسل مؤقتة أو زمنية). تُعَدُّ بعضُ طرق التحليلات التنبؤية (مثل: التنقيب في البيانات) وخوارزميات تعلم الآلة؛ طرقاً انتقائيةً للغاية لنوع البيانات التي يمكِنها التعامل معها.

وقد يــؤدي الاعتمادُ على أنواع بيانات غير متوافقة مع طــرق التحليلات إلى الحصول على نماذج غير صحيحة أو (في أغلب الأحيان) تُعَثِّر عملية تطوير النموذج. فعلى ســبيل المثال: تحتاج بعضُ طرق التنقيب في البيانات أن يتمَّ تمثيلُ جميع المتغيرات (سواء مدخلات أو مخرجات) كمتغيرات ذات قيمة رقمية (مثل: الشــبكات العصبية، آلة المتجهات الداعمة، الانحدار اللوجستي). ويتمُّ تحويلُ المتغيرات الاسمية أو الترتيبية إلى تمثيلات رقمية باستخدام أحد أنواع المتغيرات الزائفة من ١ إلى ن (فمثلًا، يمكِن تحويلُ متغيرٍ فئويٍّ له ثلاث قيم مختلفة إلى ثلاثةِ متغيراتٍ زائفة ذات قيم ثنائية ٠ أو ١). ولأن هذه العملية قد تزيدُ من عددِ المتغيرات؛ فلا بد من الحَذر بشأن تأثير هذه التمثيلات؛ خاصةً فيما يتعلق بالمتغيرات الفئوية التي تحتوي على أعدادٍ كبيرة من القيم الفريدة.

وبالمثل؛ تحتاج بعض طرق التحليلات التنبؤية، مثل ID3 (خوارزمية شجرة القرارات الكلاسيكية) والمجموعات الخام (خوارزمية قاعدة الاستقراء الجديدة نسبيًّا)، إلى تمثيلِ جميع المتغيرات كمتغيراتٍ قطعيــة القيمــة. وقد تطلَّبَت الإصــداراتُ القديمة من هذه الطرق من المســتَخدِم أن يقوم بفصلِ وتمثيــلِ المتغيِّرات الرقميــة في مجموعات فئوية قبل أن تتمَّ معالجتها بواســطة الخوارزمية. والخبر السَّــار؛ هو أنَّ معظم تطبيقات هذه الخوارزميات في أدوات البرامج المتاحة على نطاقٍ واســعٍ تقبل مزيجًا من المتغيرات الرقمية والاسمية وتقوم داخليًّا بإجراء التحويلات اللازمة قبل معالجة البيانات.

تَظهر البياناتُ في العديد من أنواع المتغيرات المختلفة والرسومات البيانية. تتحسَّن أدواتُ تحليل الأعمال باستمرار في قدرتها على مساعدةِ علماء البيانات في المهمة الشاقة المتمثلة في تحويل وتمثيل البيانات؛ بحيث يمكِن تنفيذُ متطلبات البيانات الخاصة بنماذج وخوارزميات تنبؤية مُحدَّدة بشــكلٍ صحيح. وتوضِّح الحالة العملية ٢-١ سيناريو العمل الذي قامت فيه شركةُ الأبحاث والتطوير الخاصة بالأجهزة الطبية الغنية بالبيانات بتبســيطِ أنشــطةِ التحليلات لديها؛ بهدف الوصول بسهولة إلى كلٍّ من البيانات والتحليلات التي تحتاجها لمواصلة أعمال الابتكار والجودة بها على أعلى المستويات.

حالة عملية ٢-١

شركة أجهزة طبية تضمَن جودة المنتج مع توفير المال

تتطــوَّر بعضُ التقنيات بشــكلٍ أسرع منها في المجال الطبي؛ لــذا فقد يغيِّر امتلاكُ برامــج التحليلات المتقدِّمة المناســبة من قواعد اللعبة. ويُعَدُّ معمــلُ الأجهزة رائدًا في تطويــر وتصنيع وتوزيع الأجهزة الطبية والتقنيات المتعلقة بها، بما في ذلك التقنية التي أحدثت ثورةً في اختبارات الدم الكامل والتخثر. وللمســاعدة في ضمان اســتمرار نموِّها ونجاحها، تعتمدُ الشركة على تحليلات البيانات وDell Statistica.

المشكلة:

وباعتبارها شركةً رائدةً في سوق أدوات التشخيص الخاصة بالرعاية الحرجة والتخثر؛ لذا يجبُ أن يستفيد معملُ الأجهزة من التقنيات السريعة التطور مع الحفاظ على كلٍّ مــن الجودة والكفاءة في تطوير وتصنيع وتوزيــع منتجاته. وعلى وجه الخصوص؛ فقد مكَّنَت الشركةُ علماءها ومهندســيها في مجالي البحث والتطوير من ســهولة الوصول إلى ثــروة غزيرة من البيانات ومن ثم تحليلهــا، بجانب الكفاءة في مراقبة عمليات التصنيعِ وسلاسل التوريد.

يوضِّــح جون يونغ، محلل الأعمال لمعمل الأجهزة: «مثــل العديد من الشركات، كنا أغنياء بالبيانــات؛ ولكن فقراء في التحليل؛ فلم يَعُد من الممكــن أن ينجح مُحللو البحث والتطوير في الوصــول إلى تقنيةِ المعلومات في كلِّ مرة يحتاجون فيهــا إلى الوصول إلى بيانات الاختبار، ومن ثــم إجراء تحليلات لمرةٍ واحدة في Minitab؛ بل يجبُ أن يكون لديهم القدرة على الوصول إلى البيانات بسرعةٍ وإجراء تحليلات مُعقَّدة بشكلٍ متسق ودقيق».

كان تنفيذُ تحليلات متطوِّرة لمعمل الأجهزة أمرًا حسَّاسًا؛ وذلك بسبب حجم وتعقيد منتجاتها. فعلى ســبيل المثال: تقوم الشركــة كلَّ عام بتصنيع مئات الآلاف من اللفائف التــي تحتوي على بطاقة بها مجموعة متنوعة من أجهزة الاستشــعار التي تقوم بقياس الإشارات الكهربائية للدم في أثناء عملية الاختبار.

يقــول يونج: "تتأثر أجهزةُ الاستشــعار هــذه بعوامل عديدة جــدًّا، من التغيرات البيئية، مثل: الحرارة والرطوبة إلى عدم تناســق المواد الواردة؛ لذلك فإننا نقوم وبشكلٍ متواصــلٍ بمراقبة أدائها. إننا نجمعُ ملايين السِّــجلات من البيانــات، والتي يتمُّ تخزين معظمهــا في قواعد بيانات SQL Server. وكنــا في حاجةٍ إلى مِنصة تحليلية تمكِّن فِرق البحث والتطوير التابعة لنا من الوصول بسرعة إلى تلك البيانات والتدخُّل السريع لحلِّ أيِّ مشكلات قد تحدث. وأيضًا؛ ونظرًا لوجود العديد من العوامل في اللعبة؛ فقد احتجنا إلى برنامــج تلقائي يمكنه مراقبــة بيانات الاختبار بذكاء وتنبيهنا إلى المشــكلات الطارئة بشكلٍ فوري".

الحل:

بــدأ معملُ الأجهزة بالبحث عن حلٍّ تحليــليٍّ لتلبية احتياجاته. فقامت الشركة سريعاً بإزالة معظم الأدوات من السوق؛ لأنها فشلت في تقديم الوظائف الإحصائية ومستوى الثقة

المطلوبة لبيئة الرعاية الصحية. مع ترك اثنين من المنافسين: الأول Statistica، وحل تحليلي آخر وهو Dell Statistica. بالنسبة لمعمل الأجهزة، كان الفائز الواضح هو Statistica.

يقول يونغ: «كان اختيار Statistica قرارًا سهلًا؛ فمع Statistica تمكّنا من إنشاء مجموعةٍ واسعةٍ من تشكيلات التحليل العالية المستوى للبيانات الخاصة بنا؛ وذلك ليستخدمها المحلّلون على نطاقِ المؤسسة بأكملها. والآن؛ عندما يريدون فهم أشياء مُحدَّدة؛ يمكنهم ببساطة إجراءُ تحليلٍ مُعلَّبٍ من هذا المتجر المركزي بدلًا من الاضطرار إلى طلب تقنية المعلومات للوصول إلى البيانات أو تذكر كيفية إجراء اختبارٍ معين».

إضافةً إلى ذلك؛ كانت Statistica أسهلُ بكثيرٍ في الاستخدام والنشر من حلولِ التحليلات القديمة. يُدوِّن يونج في ملاحظاته: «لتنفيذ والإبقاء على حلولٍ تحليلية أخرى، تحتاج إلى معرفةٍ برمجة حلول التحليلات؛ ولكن مع Statistica يمكنني الاتصالُ بالبيانات الخاصة بنا، وتحليلها ونشرها في غضون ساعة - على الرغم من أنني لستُ مبرمجًا عظيمًا».

وأخيرًا؛ بالإضافة إلى وظائفها المُتقدِّمة وسهولة استخدامها؛ فقد قدَّمت شركة Statistica دعمًا عالميًّا وخدمةً سعريةً جذَّابةً. يقول يونغ: «إنَّ الأشخاصَ الذين ساعدونا في تطبيق نظام Statistica كانوا ببساطة رائعين، وكان السعرُ أقلَّ بكثيرٍ مما يتيحه أيُّ حلٍّ آخر للتحليلات».

النتائج:

مع Statistica؛ أصبح الآن لدى مُحلِّلي المشروع القدرة على الوصولِ السَّهل إلى كلٍّ من البيانات والتحليلات التي يحتاجونها لمواصلة كلٍّ من الابتكار والجودة في معمل الأجهزة. في الواقع؛ فإنَّ التحليلات السريعة والفعَّالة وكذلك التحذيراتُ الآلية من جانب Statistica توفِّر للشركة مئات الآلاف من الدولارات.

يقول يونج: "في أثناء تصنيع خرطوشة الحِبر؛ نتعرَّض من حينٍ لآخر لبعض المشكلات، مثل عدم الدقة في التركيب الكيميائي الذي يمرُّ على أحدِ أجهزة الاستشعار، إن تخريد دفعة واحدة من البطاقات قد يكلِّفنا مئات الآلاف من الدولارات. غير أنَّ Statistica تساعدنا على اكتشاف الخطأ بسرعةٍ وإصلاحه مما يُجنِّبنا تلك التكاليف. فمثلًا؛ يمكننا تزويجُ بيانات الاختبار مع بيانات سجل تاريخ الأجهزة الإلكترونية من بيئة SAP الخاصة بنا وإجراء كلِّ أنواع الارتباطات لتحديد أيُّ التغيُّرات يؤدي إلى مشكلة، من هذه التغيرات مثلًا التغيُّرات في درجة الحرارة والرطوبة.

تُعَـدُّ عملياتُ التحقق من الجودة اليدوية ذات قيمة كبيرة بكلِّ تأكيد؛ ولكن تدير Statistica تلقائياً مجموعةً متنوعةً من التحليلات للشركة، بشـكلٍ يضمنُ عدمَ حدوث أيِّ خطأ واكتشاف المشكلات بسرعة. يقول يونغ: «من المقرَّر إجراء العديدِ من عمليات التحليل بشـكلٍ دوريٍّ لفحص أشياء مختلفة؛ فإذا تمَّ اكتشافُ مشكلة؛ فسيقوم النظام تلقائيًّا بتنبيه المسؤولين عبر البريد الإلكتروني أو تسجيل المخالفات بقاعدة البيانات».

إنَّ بعض المزايا الرئيسية لتحليلات البيانات المتقدِّمة مع Statistica تشتملُ على ما يلي:

- **الالتـزام بالتنظيـم:** بالإضافـة إلى توفير أمـوالِ معامل أجهزة القيـاس؛ تضمَن Statistica كذلـك التزامَ عمليات الشركة بقوانـين إدارة الأغذية والأدوية (FDA) فيما يتعلَّق بالجودة والاتساق. يُوضِّح يونج ما يلي: «لأننا نقوم بتصنيع الأجهزة الطبية؛ فإننا نخضعُ لنظام إدارة FDA، تسـاعدنا Statistica على إجراء عمليات التحقق الإحصائية التي تتطلبها إدارة FDA - فعلى سـبيل المثال: يمكننا بسهولةٍ إثباتُ أنَّ مجموعتين من المنتجات المصنوعة باستخدام مواد كيميائية مختلفة هي نفسها إحصائيًّا».

- **ضـمان الاتسـاق:** إنَّ ترتيبـات التحليـل الموحَّـدة في Statistica والتـي يمكن اسـتخدامُها في أثنـاء أيِّ مشروع تضمَن التناسـق والجودة في معمـل الأجهزة. يقول يونـج: "وعنـد تحليلِ البيانات يمكنك الحصولُ على نتائـج مختلفة تبعًا للطريقة التي تستخدمها في التحليل. فمثلًا قد يستخدم علماءُ مختلفون أنظمةً مختلفةً على البيانات، أو لا يقومون بتلخيصها إطلاقًا، عندئذٍ فسـيحصلون جميعًا على نتائج مختلفة، أمَّا مع Statistica؛ فنسـتطيع أن نضمَن أنَّ جميع العلـماء في جميع مراحل المشروع يقومون بإجراء التحليلات بنفس الطريقة، وبهذا نحصل على نتائج متناسقة.»

- **مراقبـة سلسـلة التوريد:** لا يقوم معمـلُ الأجهزة بتصنيع البطاقـة، وبها أجهزة استشـعار فحسـب؛ بل تقوم بتصنيع الجهاز الطبي بالكامـل، وبالتالي فهو يعتمدُ على الموردِّين لإمداده بالأجزاء المطلوبة. ولمزيدٍ من ضمان الجودة، تُخطِّط الشركة لتوسـيع نطاق تعاملها مع Statistica لمراقبة سلسلة التوريد.

- **توفـير الوقت:** بالإضافة إلى توفير الأموال ورفع مسـتوى الالتزام بالتنظيم لمعمل الأجهزة، تعمل شركة Statistica أيضًا على توفير الوقت الثمين لمهندسي وعلماء الشركة؛ مما يمكِّنهم من التركيز بشـكلٍ أكبر على الابتكار والحَدِّ من الروتينيات. يدوِّن يونج في ملاحظاته: «توفر التنبيهات الاسـتباقية من Statistica للمهندسـين الكثيرَ من الوقت؛ فهم ليسـوا بحاجةٍ لتذكُّر فحص العوامـل المختلفة طوال الوقت، مثل: نقص الجلوكوز؛

فهــذا اختبارٌ واحدٌ؛ لكنه يســغرق نصف اليوم. أما في ظـل مراقبة Statistica لبيانات الاختبار لدينا؛ فبإمكان المهندسـين التركيزُ على أمورٍ أخرى؛ لعلمهم بأنه سيصلهم بريدٌ إلكتروني إذا أدَّى نقصُ الجلوكوز لأيِّ مشكلة».

الاحتمالات المستقبلية:

إنَّ معمـلَ أجهـزة القياس مُتحمِّس جدًا للفرص التي تحقَّقَـت من خلال الرؤية التي يوفرهـا برنامجُ التحليلات المتقدِّمة مـن Statistica في مخازن البيانات الخاصة به. يقول يونج: "باستخدام Statistica؛ يمكنك اكتشافُ جميع أنواع الأفكار المتعلقة ببياناتك، والتي قد لا تتمَكَّن من العثور عليها. فقد يكون أمامك مكاسبُ مادية خفية؛ لكنك لا تراها لأنك لم تحلل بياناتك كما ينبغي. فباسـتخدام تلك الأداة، اكتشفنا بعض الأشياء المثيرة للاهتمام في بياناتنا والتي وفرت لنا مبلغًا كبيرًا جدًا من المال، ونتطلَّع إلى إيجاد المزيد منها».

أسئلة للمناقشة:

١- مـا هي التحدِّيات الرئيسـية لشركة الأجهـزة الطبية؟ وهل تمَّ قيادة السـوق أو التقنية؟ اشرح.

٢- ما هو الحلُّ المقترح؟

٣- ما هي النتائج التي تمَّ التوصل إليها؟ وما رأيك في العائد الحقيقي للاستثمار (ROI)؟

Source: Dell customer case study. Medical device company ensures product quality while saving hundreds of thousands of dollars. https://software.dell.com/documents/instrumentation-laboratory-medical-device-companyensures-product-quality-whilesaving-hundreds-ofthousands-of-dollars-case-study-80048.pdf (accessed August 2016). Used by Permission from Dell.

أسئلة مراجعة على القسم ٢-٣:

١- ما هي البيانات؟ وكيف تختلفُ البيانات عن المعلومات والمعرفة؟

٢- ما هي الفئاتُ الرئيسية للبيانات؟ ما هي أنواعُ البيانات التي يُمكِن استخدامها في ذكاء الأعمال والتحليلات؟

٣- هل يمكننا استخدامُ نفس تمثيل البيانات لجميع نماذج التحليلات؟ لماذا أو لماذا لا؟

٤- ما هو تمثيل بيانات ١- من - ن؟ لماذا وأين يتمُّ استخدامها في التحليلات؟

٢-٤ فن وعلم معالجة البيانات:

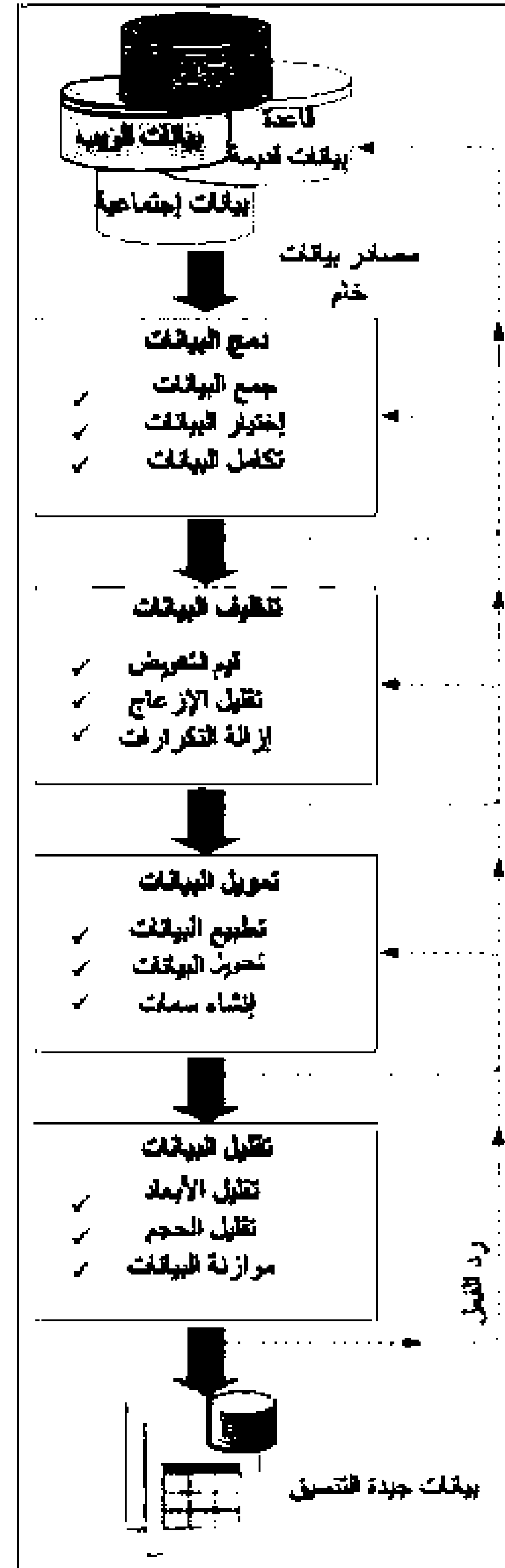
عادةً ما تكون البيانات في شكلها الأصلي (أيْ: بيانات العالم الحقيقي) ليست جاهزةً للاستخدام في مهامِّ التحليلات. فغالبًا ما تكون كثيرة الأخطاء ومنحرفةً ومُعقدةً للغاية، وغير دقيقة؛ لذا فإنه من الضروري القيامُ هنا بعملية شاقة ومُكلِّفة من ناحيةِ الوقت (تُسمَّى عملية المعالجة المسبقة للبيانات) بغرض تحويلِ بيانات العالم الحقيقي إلى شــكلٍ جيد الدقة لاســتخدامها في خوارزميات التحليــلات (Kotsiantis, Kanellopoulos, & Pintelas, 2006). ويقــرُّ العديــدُ من مُحترفي التحليلات أنَّ الوقت المُســتَغرَق في معالجة البيانات (والذي قد يكون أقلَّ مراحل العملية امتاعًا) يكــون أطول بكثيرٍ من الوقت المُســتَغرَق في بقية مهامِّ التحليلات (متعــة بناء نموذج التحليلات وتقييمها). ويوضِّح الشكل ٢-٣ الخطوات الرئيسية في محاولة معالجة البيانات.

في المرحلــة الأولى من المعالجة المُســبَقة للبيانات؛ يتمُّ جمعُ البيانات المتعلِّقة بالدراســة من المصادر المُحدَّدة لذلك، ويتمُّ تحديدُ الســجلات والمتغيرات الضرورية (استنادًا على الفهم العميق للبيانــات، مع تصفيــة المعلومات غير الضرورية)، كــما يتمُّ تكامل/ دَمْج السِّــجلات الواردة من بيانات متعددة المصادر (وأيضًا باســتخدام الفهم العميق للبيانــات يُمكِن التعامل مع المرادفات والمتضادات المتشابهة بشكلٍ صحيح).

في المرحلــة الثانية مــن معالجة البيانات؛ يتــمُّ تنظيفُ البيانات (تُعرَف هــذه الخطوة أيضاً بـ تنقيــة البيانات). فعادةً ما تكون البياناتُ في شــكلها الأصلي/ الخــام/ الحقيقي كثيرة الأخطاء (Hernández و1998 ;Stolfo, Kim وآخــرون، ٢٠٠٣). في هــذه الخطوة؛ يتمُّ تحديدُ القيم في مجموعة البيانات والتعامل معها. في بعض الحالات؛ تُشــكِّل القيمُ المفقودة شــذوذاً في مجموعة البيانــات، وفي هــذه الحالة لابدَّ من حصرهــا (بوَضْع قيم محتملة لهــا) أو تجاهلها، وفي حالاتٍ أخــرى؛ تكون القيمُ المفقودة جزءًا طبيعيًّا من مجموعة البيانــات (فمثلًا، غالبًا ما يتمُّ ترك خانة دخل الأسرة بــدون إجابة من أصحاب الدخل المرتفع). في هذه الخطوة؛ يجب أيضًا على المحلِّل تحديدُ القيم الصاخبة في البيانات (أيْ: القيم المتطرفة) ومن ثم تسهيلها. بالإضافة إلى ذلك؛ يجب التعامــلُ مع حالات عدم الاتســاق في البيانات (كوجود قيم غير عادية ضمن متغير) باســتخدام معرفة مجال الدراسة و/ أو رأي الخبير.

في المرحلــة الثالثة من معالجة البيانات؛ يتــمُّ تحويلُ البيانات من أجل معالجة أفضل. فعلى ســبيل المثال: في العديدِ من الحالات، يتمُّ توزيعُ البيانات توزيعًا طبيعيًّا بين حَدَّين (أدنى وأقصى) مُحدَّديــن؛ وذلك لجميع المتغيِّرات؛ بهدف تخفيف التحيُّــز المحتمل لمتغيرٍ واحدٍ (به قِيَم عددية

كبـيرة، مثل دخـل الأسرة) يحتوي على متغيرات أخـرى (لها قيمٌ أصغر، مثل عـدد مَنْ يعول أو سـنوات الخدمـة) والتي قد تكون أكثر أهميةً. وهناك تحويلٌ آخـر يتمُّ إجراؤه وهو التفريد و/ أو التجميـع. وفي بعض الحالات؛ يتـم تحويلُ المتغيرات الرقمية إلى قيم فئوية (مثل: منخفض - متوسط - مرتفع)، وفي حالاتٍ أخرى؛ يتمُّ تقليلُ النطاق الفريد لقيمة المتغير الاسمي إلى مجموعةٍ أصغر باسـتخدام مفهومِ التسلسل الهرمي (على سبيل المثال: عند استخدام حالات فردية تحتوي على ٥٠ قيمة مختلفة، قد يستخدمُ المحلِّل عِدَّة مناطق لمتغيرٍ يُظهر الموقع)؛ وذلك للحصول على مجموعةِ بيانات تكون أكثر قابليةً للتعامل مع الحاسب. ومع ذلك؛ ففي حالاتٍ أخرى، قد يختار المحلِّل إنشـاءَ متغيرات جديدة تعتمدُ على المتغيرات الحالية لزيادة معنوية المعلومات الموجودة بمجموعة المتغيرات في حزمة البيانات.



شكل ٢-٣: خطوات معالجة البيانات

فعلى سـبيل المثال: في مجموعةِ البيانات الخاصة بزراعة الأعضاء، قد يختارُ المحلل اسـتخدامَ متغـيرٍ مفـردٍ يوضِّح حالة الدم (1 = match, 0 = no-match) بدلًا من اسـتخدام قيمٍ منفصلةٍ متعددة الصور لنوع الفصيلة لدى كلٍّ من المتبرع والمتلقي. وقد يؤدي هذا التبسـيط إلى زيادة محتوى المعلومات مع تقليل تعقيد العلاقات في البيانات.

أمَّـا المرحلةُ النهائية مـن معالجة البيانات الأوَّلية فهي تقليـل البيانات. وعلى الرغم من أن علـماء البيانات (أيْ: محترفو التحليلات) يحبُّون الحصول على مجموعاتٍ كبيرة من البيانات؛ فإنَّ البيانات الزائدة قد تكون أيضًا مشكلة. بمعنى أبسط؛ يمكِن للمحلِّل أن يتصور البيانات الشائعة الاسـتخدام في مشـاريع التحليلات التنبؤية كملفٍ ثابـت يتكوَّن من بُعدَيـن: المتغيرات (عدد الأعمدة) والحالات/ السـجلات (عدد الصفوف). في بعض الحالات (كمعالجة الصور ومشـاريع genome مـع بيانـات microarray معقدة)، يمكـن أن يكون عددُ المتغـيرات (التي تمَّ تمثيلها بالأعمدة) كبيرًا نسـبيًّا، ويجب على المحلِّل تقليل العدد إلى حجمٍ يسـهل التحكُّم فيه. تُسـمَّى هذه العملية عـادةً تخفيض الأبعاد (أو انتقاء المتغير)؛ وذلك لأن المتغيرات تتمُّ معالجتها كأبعاد مختلفة تَصِف ظاهرةً من منظورات مختلفة، في التحليلات التنبؤية والتنقيب في البيانات. وعلى الرغم من عدم وجود طريقة واحدة مُثلى لإنجاز هذه المهمة؛ فإنه يمكِن للمحلل استخدامُ نتائج سـابقة النشر، واستشـارة خبراء المجال محل الدراسـة، وإجراء اختبارات إحصائية مناسبة (مثل: تحليل المكوِّنات الأساسـية أو تحليل المكوِّنات المسـتقلة)، والأكثر تفضيلًا هو استخدام مزيجٍ من هذه التقنيات لتقليلِ أبعاد البيانات بنجاح إلى مجموعةٍ فرعيةٍ أكثر قابليةً للإدارة وأكثر ملاءمةً.

وفيـما يتعلَّـق بالبُعْد الآخر (أيْ: عـدد الحالات والتي يتمُّ تمثيلهـا في الصفوف)؛ قد تتضمَّن بعضُ حُزم البيانات ملايين أو مليارات من السِّجلات. وعلى الرغم من ازدياد قوة الحوسبة بصورةٍ مضاعفة؛ فإنَّ معالجة مثل هذا العدد الكبير من السجلات قد لا يكون عمليًّا أو ممكنًا. وفي مثلِ هـذه الحـالات؛ قد نحتاجُ إلى تحليل عينةٍ عبارة عن حزمة فرعية مـن البيانات. ويُعَدُّ الافتراض الأسـاسي لأخذ العينات هو أنَّ المجموعة الفرعية من البيانات سـوف تحتوي على جميع الأنماط المتعلقة بمجموعة البيانات الكاملة. يجب أن يكون المحللُ شديدَ الحذر عند اختياره للمجموعة الفرعيـة من البيانات، والتي يجبُ أن تعكسَ جوهر مجموعة البيانات الكاملة، وليسـت خاصةً بمجموعـةٍ فرعية أو فئة فرعية. ويتمُّ فرزُ البيانات عادةً عـلى بعض المتغيِّرات، وقد يقودنا أخذُ جـزءٍ مـن أعلى البيانات أو أسـفلها إلى بيانات مُتحيِّزة لقيم مُحدَّدة مـن المتغير المفهرس؛ لذلك حاول دائمًا اختيار السِّـجلات الموجودة في مجموعة العينات بشـكلٍ عشوائي. وبالنسبة للبيانات

المنحرفة، قد لا يكون أخذُ عيناتٍ عشوائية بصورة مباشرة كافيًا؛ بل قد يكون الأفضلُ هو أخذ عينةٍ طبقية (أيْ: إجراء تمثيلٍ نسـبي لمجموعاتٍ فرعية مختلفة مـن البيانات المتاحة ضِمن مجموعة بيانـات العينة). وعنـد التعامل مع البيانات المنحرفـة فحريٌّ بنا تحقيق التـوازن بين البيانات شـديدة الانحـراف؛ إما من خلال الإفراط أو التقصير في تمثيل الفئـات الأكثر تمثيلًا. وقد أظهرتِ الأبحاثُ أنَّ مجموعات البيانات المتوازنة تميل إلى إنتاج نماذج تنبؤ أفضل من النماذج غير المتوازنة (Thammasiri وآخرون، ٢٠١٤م).

يتمُّ تلخيصُ جوهر معالجة البيانات في الجدول ٢-١، الذي يرسـم المراحل الرئيسـة (جنبًا إلى جنب مع أو صاف مشكلاتها) إلى قائمة تمثيلية للمهامِّ والخوارزميات.

يكاد يكون من المسـتحيل التقليلُ من قيمة اقتراح معالجـة البيانات. وهي واحدةٌ من تلك الأنشـطة التي تتطلب وقتًا طويلًا؛ حيث يؤتي الاستثمار في الوقت والجهد ثمارًا هائلةً تغطي على تناقص العائدات. وهذا يعني أنه كلما زادت الموارد التي تسـتثمر فيها؛ زادت الأرباح في النهاية. توضِّـح الحالة العملية ٢-٢ دراسـةً مثيرةً للاهتمام؛ إذ يتمُّ اسـتخدام البيانـات الأكاديمية الخام والمتاحـة بسـهولةٍ داخل منظمة تعليميـة لتطوير نماذج تنبؤية لفهم تـسرُّب الطلاب المبتدئين، وتحسـين اسـتبقاء الطلاب الجدد في مؤسسـة التعليم العالي الكبيرة. بناءً على ما هو مُوضَّح في الحالة العملية؛ فإنَّ كلَّ مهمةٍ من عمليات معالجة البيانات الموضَّحة في الجدول ٢-١ كانت مهمةً للتنفيذ الناجح لمشروع التحليلات الأساسية، لا سيما المهمة المتعلقة بموازنة مجموعة البيانات.

جدول ٢-١ : ملخص لمهامِّ معالجة البيانات، والأساليب المحتملة

المهمة الرئيسية	المهام الفرعية	الطرق الشائعة
توحيد البيانات	- الوصول للبيانات وجمعها. - اختيار وتنقية البيانات. - تكامل وتوحيد البيانات.	- اسـتعلامات SQL، وكلاء البرامج، خدمات الويب. - الخـبرة في المجـال، اسـتعلامات SQL، الاختبارات الإحصائية. - اسـتعلامات SQL، وخبرة المجال، وتصنيف البيانات المعتمدة على علم الأنتولوجي.

المهمة الرئيسية	المهام الفرعية	الطرق الشائعة
تنظيف البيانات	- التعامل مع القِيم المفقودة في البيانات.	- ملء القيم المفقودة (الانحرافات) مع القيم الأكثر ملاءمةً (الوسيط، الوسط الحسابي، الحدُّ الأدنى/ الحدُّ الأقصى، المنوال... إلخ). - إسناد قيمةٍ ثابتةٍ، مثل "ML" للقيم المفقودة؛ إزالة سجل القيمة المفقودة؛ عدم فعل أيِّ شيء.
	تحديد وتقليل الضوضاء في البيانات.	- تحديد القيم المتطرفة في البيانات باستخدام تقنيات إحصائية بسيطة (مثل: المتوسطات، والانحرافات المعيارية) أو بالتحليل العنقودي؛ أو بمجرد تحديدها نقوم بإزالة القيم المتطرفة أو تسهيلها باستخدام الثنائيات أو الانحدار أو المتوسطات البسيطة.
	إيجاد، وإزالة البيانات الخطأ.	- تحديد القِيم الخطأ في البيانات (بخلاف القيم المتطرفة)، مثل: القيم الفردية، تصنيفات الفئات غير المتسقة، التوزيعات الفردية؛ وبمجرد تحديدها، استخدم خبرات المجال لتصحيح القيم الخطأ، أو إزالة السجلات المشتملة على قيمٍ خطأ.
تحويل البيانات	توزيع البيانات طبيعيًّا.	- تقليل نطاق القِيم في كلِّ متغيرٍ ذي قيمة رقمية إلى نطاق معياري (على سبيل المثال: ٠ إلى ١ أو -١ إلى +١) باستخدام مجموعةٍ متنوعةٍ من تقنيات التورزيع الطبيعي أو القياس.
	فصل أو تجميع البيانات.	- إذا لزم الأمر؛ نقوم بتحويل المتغيرات الرقمية إلى تمثيلات منفصلة باستخدام تقنيات متعلقة بنطاق أو تكرار المتغيرات الفئوية، والحَدّ من عدد القيم من خلال تطبيق التسلسل الهرمي للمفاهيم السليمة.

المهمة الرئيسية	المهام الفرعية	الطرق الشائعة
	بناء سماتٍ جديدة.	- استنباط مُتغيرات جديدة وذات معلومات أكثر من المتغيّرات الموجودة بالفعل باستخدام مجموعةٍ واسعةٍ من الدوال الرياضية (بسيطة مثل الجمع والضرب، أو معقدة مثل مزيجٍ مختلط من تحوُّلات السجل).
تقليل المعلومات	تقليل عددِ السِّمات.	- تحليل المكوِّن الرئيسي، تحليل مكوِّن مستقل، اختبار chi-square، تحليل الارتباط، واستقراء شجرة القرارات.
	تقليل عدد السِّجلات.	سحب عشوائي للعينات، سحب طبقي للعينات، سحب متأنٍّ للعينات باستخدام الخبرة والمعرفة السابقة.
	توازن البيانات المنحرفة.	استخلص أقلَّ تمثيل أو تمثيل الفئات الأكثر تمثيلًا

حالة عملية ٢-٢

تحسين الاحتفاظ بالطلاب باستخدام التحليلات المعتمدة على البيانات

أصبحت ظاهرةُ تسرُّب الطلاب واحدةً من أكثر التحَدِّيات التي تواجه صانعي القرار في المؤسسات الأكاديمية. فعلى الرَّغم من كلِّ البرامج والخدمات التي تمَّ وضعها للمساعدة في الاحتفاظ بالطلاب؛ وفقًا لوزارة التعليم الأمريكية ومركز الإحصائيات التعليمية (nces.ed.gov)؛ فلا يحصل على درجة البكالوريوس أكثر من نصف الذين يلتحقون بالتعليم العالي. أصبحت إدارة التسجيل والإبقاء على الطلاب أولويةً قصوى لمديري الكليات والجامعات في الولايات المتحدة وبلدان أخرى حول العالم. فعادةً ما يؤدي التسرُّب الكبير للطلاب إلى خسارة مالية عامة، وانخفاض معدلات التخرُّج، وسمعة تربوية أدنى في نظر جميع أصحاب المصلحة. إذ يبحث كلٌّ من المشرِّعين وصانعي السياسات الذين يشرفون على التعليم العالي ويُخصِّصون له الموازات التقديرية، وكذلك الآباءُ الذين يدفعون تكاليف تعليم أبنائهم؛ بغرض إعدادهم لمستقبلٍ أفضل، وأيضًا الطلاب الذين يختارون التخصُّصات الجامعية - يبحث كلُّ هؤلاء - عن أدلة على الجودة المؤسسية والسُّمعة لتوجيه عمليات صُنع القرار الخاصة بهم.

الحلُّ المقترَح:

لتحســين الاحتفــاظ بالطلاب؛ يجب عــلى المحلِّل أولاً أن يحاول فهم الأســباب غير الواهيــة وراء التــسرُّب من التعليم. ولكي ننجــح في ذلك؛ علينــا أولًا أن نحَدِّد وبدقة الطــلاب الذين يُداهمهم خطر التسرُّب. حتى الآن؛ تــمَّ تكريسُ الغالبية العظمى من بحوث اســتنزاف الطــلاب لفهم هذه الظاهرة الاجتماعية المعقدة والحاســمة في نفس الوقت. وعلى الرَّغم من أن هذه الدراســات النوعية والسُّــلوكية والاستقصائية كشفت عــن رؤية ثمينة للغاية من خلال تطوير واختبار مجموعة واســعة من النظريات؛ فإنها لا توفــر الآلية اللازمة للتنبؤ الدقيق (والتطوُّر المحتمل) لتسرُّب الطلاب. قدَّم المشروع المختصر في دراســة هذه الحالة نهجًا بحثيًّا كميًّا؛ إذ يمكن اســتخدام البيانات المؤسسية التاريخية من قواعد بيانات الطلاب لتطوير نماذج قادرة على التنبؤ بالإضافة إلى توضيح طبيعة المؤسسة الخاصة بمشكلة التسرُّب. ويوضِّح الشكل ٢-٤ منهج التحليلات المقترح.

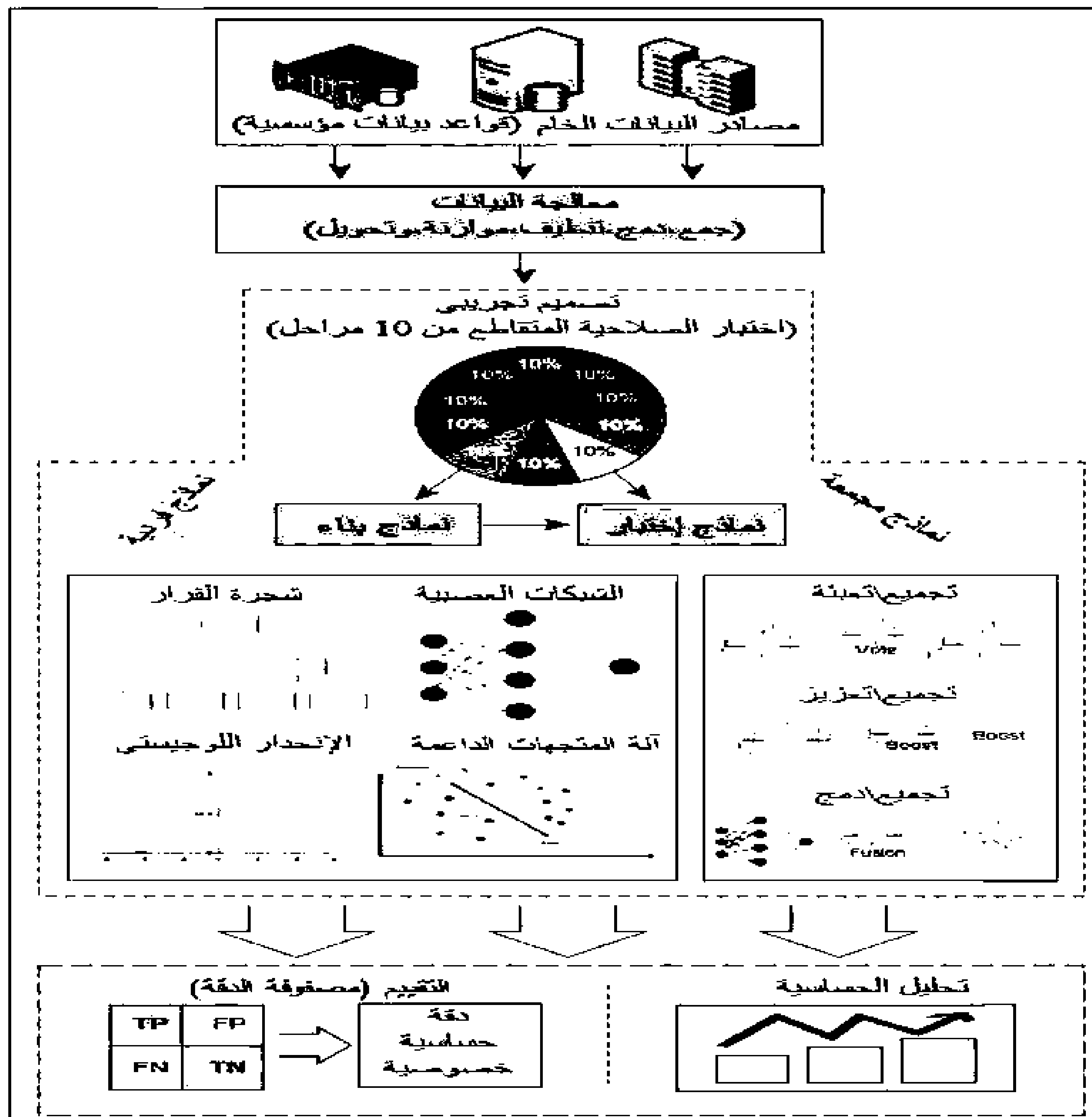
وعلى الرغم من كون المفهوم جديداً نســبيًّا عــلى التعليم العالي لأكثر من عقد من الزمان؛ فقد تمَّت دراســة مشــكلات مماثلة في مجال إدارة التسويق باستخدام تقنيات تحليلات البيانات التنبؤية تحت اســم «churn analysis»؛ بغرض الإجابة عن الســؤال التــالي: «مَن مِن عملائنــا الحاليين أكثر ميلاً للتوقف عــن شراء منتجاتنا أو خدماتنا؟» بحيــث يمكن إجــراء نوع من عمليات الوســاطة أو التدخــل للإبقاء عليهــم؛ إذ يُعَدُّ الاحتفــاظ بالعملاء الحاليين أمراً بالــغ الأهمية؛ لأننا وكما نعلــم جميعًا، وكما أظهرت البحوث المتعلقة مرارًا وتكرارًا، أن تكاليف وجهد ووقت كَسْــب عميل جديد أكثر بقدرٍ كبيرٍ من محاولة الحِفاظ على العميل الموجود بالفعل.

البيانات من الجوهر:

جاءت بيانات هذا المشروع البحثي من مؤسســة واحدة (وهي جامعة عامة شاملة تقع في منطقة الغرب الأوســط بالولايات المتحدة) بمتوسط تسجيلٍ يبلغ ٢٣٠٠٠ طالب، منهم حوالي ٨٠٪ من المقيمين في نفس الولاية، وحوالي ١٩٪ منهم من الأقليات. لا يُوجَد فرقٌ معنوي بين الجنســين في أرقام التســجيل. كان متوسط نســبة الاحتفاظ بالطلاب الجُدد في المؤسسة حوالي ٨٠٪، ومتوسط نسبة التخرُّج لمدة ٦ سنوات حوالي ٦٠٪.

استخدمت الدراسة ٥ سنوات من البيانات المؤسسية، والتي شملت أكثر من ١٦٠٠٠ طالب مُســجلين كطلاب جُدُد، تمَّ جمعهم من قواعد بيانات مختلفة ومتنوعة للطلاب الجامعيــين. وقد تضمَّنَت البيانــاتُ المتغيرات المتعلقة بالخصائــص الأكاديمية والمالية

والتوزيع السكاني للطلاب. وبعد دمج وتحويل البيانات المتعددة الأبعاد لكلِّ طالب إلى ملف مُسطَّح واحد (أيْ: ملف يحتوي على أعمدة تمثِّل المتغيرات وصفوف تمثل سجلات الطالب)، تم تقييمُ الملف الناتج ومعالجته مسبقاً لتحديد ومعالجة القيم الشاذة وغير القابلة للاستخدام. فعلى سبيل المثال: قامت الدراسة بإزالة كل سجلات الطلاب الدوليين من حزمة البيانات؛ لأنها لم تحتوِ على معلومات عن بعض المتنبئين ذوي السُّمعة الطيبة (مثل: المدرسة الثانوية GPA، ودرجات SAT). في مرحلة تحويل البيانات؛ تم تجميعُ بعض المتغيرات، (فمثلاً: يتم تجميع المتغيرات "الرئيسية" و"الكثيفة" إلى متغيرات ثنائية MajorDeclared وConcentrationSpecified) للحصول على تفسير أفضل للنمذجة التنبؤية. بالإضافة إلى ذلك؛ تمَّ استخدام بعض المتغيرات لاشتقاق متغيرات جديدة (مثل: نسبة المكتسب/ المسجل وسنوات ما بعد المدرسة الثانوية).



شكل ٢-٤: منهج التحليلات للتنبؤ بتسرُّب الطلاب

المكتسبة/ المسجلة = الساعات المكتسبة/ الساعات المسجلة

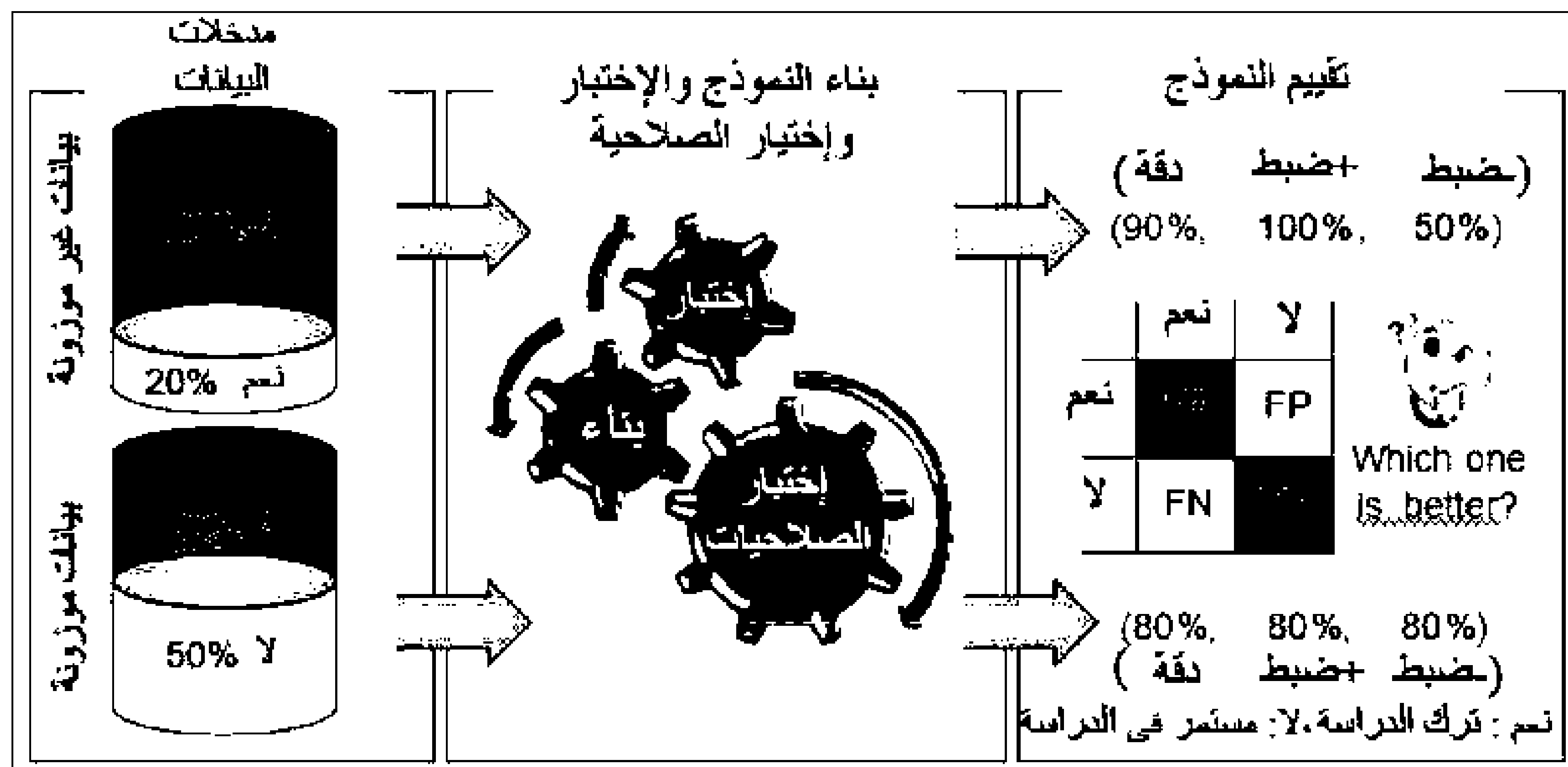
سنوات ما بعد المدرسة الثانوية = سنة تسجيل في الجامعة - سنة التخرُّج من المدرسة الثانوية

تمَّ إنشاء النسبة المكتسبة/ المسجَّلة؛ للحصول على تمثيلٍ أفضل لمرونة الطلاب وتقريرهم في الفصل الدراسي الأول من السنة الدراسية الأولى في الجامعة. وبشكلٍ منطقــي؛ فإنــه من المتوقع أن تؤثــر القيم الأكبر لهذا المتغير بشــكلٍ إيجابي على الاســتبقاء/ الثبات. تمَّ إنشــاء مصطلح سنوات ما بعد المدرســة الثانوية لقياس تأثير الوقت المًســتغرَق بين التخرُّج من المدرسة الثانوية والتسجيل الأوَّلي بالكلية. ومنطقيًّا؛ فمن المتوقع أن يُسهم هذا المتغيِّر في التنبؤ بتسرُّب الطلاب. يتم تحديدُ هذه المجموعات والمتغيِّرات المشتقة بناءً على عددٍ من التجارب التي تمَّ إجراؤها لعــدد من الفرضيــات المنطقية. وتمَّ حفظ هذه التجــارب والفرضيات في حزمة المتغير النهائي، وهي التي رفعت مســتوى الحساســية وأيضًا أدَّت إلى الوصول إلى مســتوى أفضل من دقة التنبؤ. تعكس الطبيعة الحقيقية للســكان (أيْ: الطلاب الجــدد)، والمتغير التابــع (أيْ: "Second Fall Registered") والذي احتوى على الكثير من سجلات نعم (~ ٨٠٪) وسجلات لا (~ ٢٠٪؛ انظر: الشكل ٢-٥).

تظهر البحوث أن وجود مثل هذه البيانات غير المتوازنة يؤثر ســلبيًّا على أداء النموذج. ولذلك؛ جرَّبت الدراســة خيارات اســتخدام ومقارنة نتائج نفس النوع مــن النماذج التي تمَّ بناؤها مع البيانات الأصلية غير المتوازنة (متحيِّزة لســجلات نعم) والبيانات المتوازنة جيدًا.

النمذجة والتقييم:

قامت الدراسة باستخدام أربع طرق تصنيف شائعة الاستخدام (أيْ: الشبكات العصبيــة الاصطناعية، وشــجرات القــرار، وآلة المتجهــات الداعمــة، والانحدار اللوجســتي) جنبــاً إلى جنبٍ مع ثلاثة أطقم من تقنيــات المجموعات النموذجية (أيْ: تعبئــة، وفصل، ودمج المعلومات). ثم تمَّت بعد ذلك مقارنة النتائج التي تمَّ الحصــول عليها من جميع أنواع النماذج بعضها مع بعض باســتخدام طرق تقييم نموذج التصنيف المنتظم (مثل: الدقة التنبؤية الشــاملة، الحساسية، النوعية) على عيّنات التحمُّل.



شكل ٢-٥: تصوير بياني لمشكلة عدم توازن الفئات

في خوارزميات تعلُّم الآلة (والتي سيتمُّ تغطية بعضٍ منها في الفصل ٤)، يُعَدُّ تحليل الحساسية طريقةً لتحديد علاقة "السبب والنتيجة" بين مدخلات ومُخرجات نموذج التنبؤ المعطى. وتُعَدُّ الفكرة الأساسية وراء تحليل الحساسية أنه يقيس أهمية متغيرات التوقع على أساس التغيير في أداء النمذجة الذي يحدث إذا لم يتم تضمين متغير توقع في النموذج. ويُطلَق على ممارسة النمذجة والتجارب هذه أيضًا مصطلح (a leave-one-out assessment). ومن ثم؛ فإن مقياس حساسية متغير تنبؤ مُحدَّد هو نسبة الخطأ في النموذج المدرّب بدون متغير التنبؤ إلى خطأ النموذج الذي يتضمن متغير المتنبئ هذا. وكلما كانت الشبكة أكثر حساسية لمتغير معين، زاد انخفاضُ الأداء في غياب هذا المتغير، ومن ثم زادت نسبة الأهمية. وبالإضافة إلى القوة التنبؤية للنماذج؛ فقد أجرت الدراسة أيضًا تحليلات الحساسية لتحديد الأهمية النسبية لمتغيرات الإدخال.

النتائج:

في المجموعة الأولى من التجارب، استخدمت الدراسة مجموعة البيانات الأصلية غير المتوازنة. واستنادًا إلى نتائج تقدير الصلاحية المتقاطع ذي الـ ١٠ مراحل فقد أنتجت خوارزمية آلة المتجهات الداعمة (SVM) أفضل دقة بمعدل تنبؤ إجمالي يبلغ ٨٧,٢٣٪، وجاءت شجرة القرار في المركز الثاني بمعدّل تنبؤ إجمالي قدره ٨٧,١٦٪، تليها الشبكات العصبية الاصطناعية بمعدل تنبؤ إجمالي

٨٦,٤٥٪، ثم الانحدار اللوجستي بمعدل تنبؤ إجمالي قدره ٨٦,١٢٪، (انظر: الجدول ٢-٢). ويكشف الفحص الدقيق لهذه النتائج أنَّ دقة التنبؤ لفئة "نعم" أعلى بشكل معنوي من دقة التنبؤ للفئة "لا". في الواقع؛ تنبأت جميع أنواع النماذج الأربعة بالطلاب الذين يُرجح عودتهم للسنة الثانية بدقة تبلغ ٩٠٪؛ لكنها (أيْ: النماذج) تنبأت بالطلاب المحتمل أن ينقطعوا عن الدراسة بعد عامٍ دراسي جديد بمستوى دقة لا يصل إلى ٥٠٪. ولأن التنبؤ بفئة "لا" هو الهدف الرئيسي من هذه الدراسة؛ فإن دقة أقل من ٥٠٪ لهذه الفئة تُعَدُّ غير مقبولة. ومن الممكن (بل من الواجب) أن يُعزَى مثل هذا الاختلاف في دقة التنبؤ للفئتين إلى الطبيعة غير المتوازنة لمجموعة البيانات التدريبية (أي: ~ ٨٠٪ "نعم" و~ ٢٠٪ "لا").

TABLE 2.2 Prediction results for the Original/unbalanced Data set

	ANN(MLP)		DT(C5)		SVM		LR	
	No	Yes	No	Yes	No	Yes	No	Yes
No	1494	384	1518	304	1478	255	1438	376
Yes	1596	11142	1572	11222	1612	11271	1652	11150
SUM	3090	11526	3090	11526	3090	11526	3090	11526
Per-Class Accuracy	48.35%	96.67%	49.13%	97.36%	47.83%	97.79%	46.54%	96.74%
Overall Accuracy	86.45%		87.16%		87.23%		86.12%	

استخدمت الجولة التالية من التجارب مجموعة بيانات متوازنة؛ إذ يتم تمثيل الفئتين بالتساوي تقريباً. وفي سبيل تحقيق هذا الأسلوب؛ أخذت الدراسة جميع العينات من فئة الأقلية (أيْ: الطبقة «لا») واختارت عشوائياً عدداً متساوياً من العينات من طبقة الأغلبية (أيْ: الطبقة «نعم») وتمَّ تكرارُ هذه العملية ١٠ مرات لتقليل التحيُّز المحتمل للعينة العشوائية. وتمخَّضَت كلُّ عملية من عمليات أخذ العينات هذه عن مجموعة بيانات تضمُّ ما يزيد عن ٧٠٠٠ سجل، والتي تمَّ فيها تمثيلُ كلا الفئتين («نعم» و«لا») بالتساوي تمامًا. مرةً أخرى، وباستخدام منهجية تقدير الصلاحية ذي الـ ١٠ مراحل؛ قامت الدراسة بتطوير واختبار نماذج تنبؤ لجميع أنواع النماذج الأربعة. ويبَيِّن الجدول ٢-٣ نتائج هذه التجارب. وبناءً على نتائج العينة المتحصَّل عليها من قبل؛ قامت آلة المتجهات الداعمة مرةً أخرى بإفراز أفضل معدل دقة للتنبؤ بشكلٍ عام بنسبة ٨١,١٨٪، متبوعة بأشجار القرار، والشبكات العصبية الاصطناعية، والانحدار اللوجستي بمعدلات دقة تنبؤ عامة تبلغ ٨٠,٦٥٪، و٧٩,٨٥٪، و٧٤,٢٦٪ على الترتيب. وكما يتضح من معدلات الدقة لكلِّ فئة؛ فإن نماذج التنبؤ كانت أفضل بشكلٍ معنوي في التنبؤ بفئة "لا" باستخدام

البيانات المتوازنة مقارنةً بها عند استخدام البيانات غير المتوازنة. وبشكلٍ عام؛ كانت تقنيات تعلُّم الآلة الثلاثة أفضل بكثير من نظيرتها الإحصائية الانحدار اللوجستي.

TABLE 2.3 Prediction results for the balanced Data set

Confusion Matrix	ANN(MLP)		DT(C5)		SVM		LR	
	No	Yes	No	Yes	No	Yes	No	Yes
No	2309	464	2311	417	2313	386	2125	626
Yes	781	2626	779	2673	777	2704	965	2464
SUM	3090	3090	3090	3090	3090	3090	3090	3090
Per-class Accuracy	74.72%	84.98%	74.79%	86.50%	74.85%	87.51%	68.77%	79.74%
Overall Accuracy	79.85%		80.65%		81.18%		74.26%	

بعد ذلك؛ تـمَّ إجراءُ مجموعة أخرى من التجارب لتقدير القدرة التنبؤية للنماذج الثلاثة. واستنادًا إلى نتائج تقدير الصلاحية المتقاطع ذي الـ ١٠ أجزاء، أنتج نموذج مجموعة اندماج المعلومات أفضل النتائج بمعدل تنبؤ إجمالي يبلغ ٨٢,١٠٪، ثم مجموعات من نوع التعبئة ومجموعات من نوع الدَّعم بمعدلات تنبؤ إجمالية تبلغ ٨١,٨٠٪ و٨٠,٢١٪، على التوالي (انظر: الجدول ٢-٤). وعلى الرغم من أن نتائج التنبؤات أفضل إلى حَدٍّ ما من النماذج الفردية؛ فإنه من المعروف أن مجموعات النماذج تنتج أنظمة تنبؤ أكثر متانةً مقارنةً بنموذج التنبؤ الأحادي (وسيتمُّ توضيح ذلك بمزيدٍ من التفصيل في الفصل ٤).

TABLE 2.4 Prediction results for the three ensemble Models

	Boosting (Boosted Trees)		Bagging (Random Forest)		Information Fusion (Weighted Average)	
	No	Yes	No	Yes	No	Yes
No	2242	375	2327	362	2335	351
Yes	848	2715	763	2728	755	2739
SUM	3090	3090	3090	3090	3090	3090
Per-Class Accuracy	72.56%	87.86%	75.31%	88.28%	75.57%	88.64%
Overall Accuracy	80.21%		81.80%		82.10%	

إضافةً إلى تقدير دقة التنبؤ لكلِّ نوع من أنواع النماذج؛ تمَّ إجراءُ تحليل الحساسية أيضًا باستخدام نماذج التنبؤ المطوَّرة لتحديد الأهمية النسبية للمتغيرات المستقلة (أيْ: المتنبئين). ولتحقيق نتائج تحليل الحساسية الشاملة؛ أنتج كلُّ نوع من أنواع النماذج الفردية الأربعة مقاييس حساسية خاصة به؛ بحيث تعمل على تصنيف جميع المتغيرات المستقلة في قائمة مرتبة بالأولوية. وكما هو متوقع؛ فقد أفرز كلُّ نوعٍ من أنواع النماذج مراتب حساسية متفاوتة نوعًا ما من المتغيرات المستقلة. وبعد جَمْع كلِّ المجموعات الأربع من أرقام الحساسية؛ يتمُّ توزيع أرقام الحساسية تلك توزيعًا طبيعيًّا وتجميعها ورسمها في شكلٍ بياني أفقي (كما بالشكل ٢-٦).

الاستنتاجات:

أظهرت الدراســةُ أن وجود بيانات كافية بمتغيِّرات ملائمة يقود إلى أساليب استخراج بيانات لديها القدرة على التنبؤ بنســبة تسرُّب الطلاب الجُدد بدقة تصل إلى ٨٠٪ تقريبًا. كما أظهرت النتائج أنه، وبغض النظر عن نموذج التنبؤ المُستَخدَم؛ فإن مجموعة البيانات المتوازنة (مقارنة بمجموعة البيانات غير المتوازنة/ الأصلية) أنتجت نماذج أفضل في التنبؤ بالطلاب الذين من المرجَّح تسرُّبهم من الكلية قبل سنتهم الدراسية الثانية. ومن بين نماذج التنبؤات الفردية الأربعة المُســتَخدَمة في هذه الدراســة، كان أداء آلة المتجهات الداعمة هو الأفضل، تليها أشــجار القرار، ثم الشــبكات العصبية، وأخيرًا الانحدار اللوجستي. أمّا من زاوية ســهولة الاســتخدام؛ فعلى الرغم من حقيقة أن آلة المتجهات الداعمة أظهرت نتائج أفضل للتنبؤ؛ فإنَّ القائم بالتحليل قد يختار اســتخدام أشجار القرار؛ لأنها بالمقارنة مع آلة المتجهات الداعمة والشــبكات العصبية؛ فهي تصوِّر هيكل نموذج أكثر شــفافية. أيضًا تظهر أشجار القرار بشكلٍ واضح عملية استنتاج التنبؤات المختلفة؛ مما يوفر تبريرًا منطقيًّا لنتائج معينة، في حين أن آلات ناقلات الدعم والشــبكات العصبية الاصطناعية ما هي إلا نماذج رياضية لا تقدِّم رؤية شفافة حول «كيف يفعلون ما يفعلون؟».

أسئلة للمناقشة:

١- ما هو تسرُّب الطلاب، ولماذا يُعَد مشكلة مهمة في التعليم العالي؟

٢- ماذا كانت الطرق التقليدية للتعامل مع مشكلة تسرُّب الطلاب؟

٣- قم بسَرْد ومناقشة التحَدِّيات المتعلقة بالبيانات في سياق دراسة هذه الحالة.

٤- ماذا كان الحلُّ المقترح؟ وماذا كانت النتائج؟

Sources: Thammasiri, D., Delen, D., Meesad, P., & Kasap N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. Expert Systems with Applications, 41(2), 321330-; Delen, D. (2011). Predicting student attrition with data mining methods. Journal of College Student Retention, 13(1), 1735-; Delen, D. (2010). A compara- tive analysis of machine learning techniques for student retention management. Decision Support Systems, 49(4), 498506-.

أسئلة مراجعة على القسم ٢-٤:

١- لماذا لا تكون البيانات الأصلية/ الخام قابلة للاستخدام بسهولة؛ من خلال مهامِّ التحليلات؟

٢- ما الخطوات الرئيسية لمعالجة البيانات؟

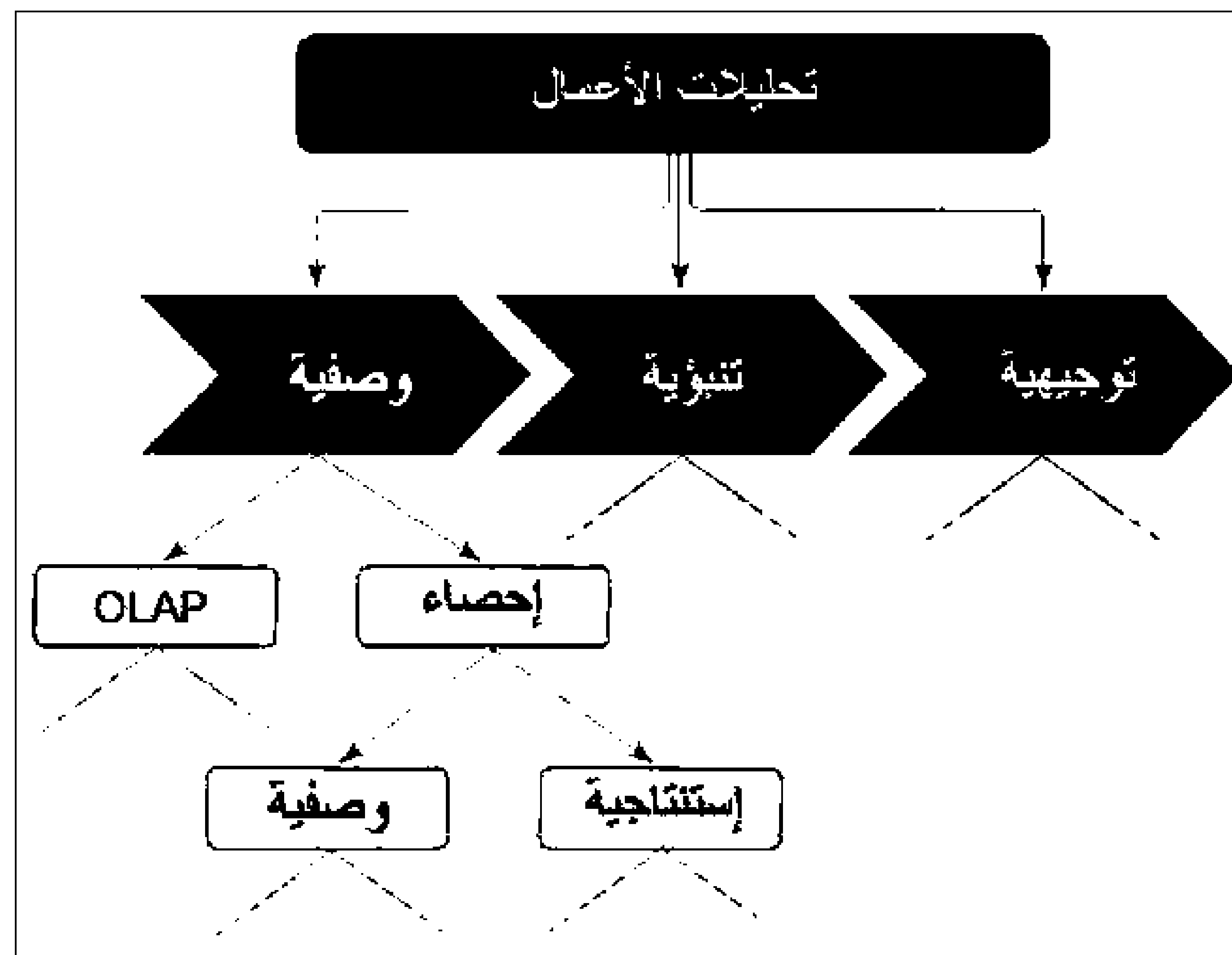
٣- ما المقصود بـ تنظيف/ تنقية البيانات؟ وما الأنشطة التي يتمُّ القيام بها في هذه المرحلة؟

٤- لماذا نحتاج إلى تحويل البيانات؟ وما هي مهام تحويل البيانات شائعة الاستخدام؟

٥- يمكن تخفيض البيانات على الصفوف (أخذ العينات) و/ أو الأعمدة (اختيار متغير). أيهما أكثر تحديًا.

٢-٥ النمذجة الإحصائية لتحليلات الأعمال:

بسـبب تزايد رواج تحليلات الأعمال؛ فإن الطرق الإحصائية التقليدية والتقنيات الأساسـية تسـتعيد أيضًا جاذبيتها كأدوات تمكينية لدعم صُنع القرارات الإدارية المبنية على الأدلة. وهي بذلك لا تسـتعيد فقط الاهتمام والإعجاب، ولكنها الآن، تجذب مسـتخدمي الأعمال إضافةً إلى خبراء الإحصـاء ومحترفي التحليلات. عادةً ما تُعتبَر الإحصائيـات (الطرق الإحصائية والتقنيات الأساسـية) جزءًا من التحليلات الوصفية (انظر: الشكل ٢-٧). كما يمكن اعتبار بعض الأساليب الإحصائية أيضًا جزءًا من التحليـلات التنبؤية مثل تحليل التمايز، والانحدار المتعدد، والانحدار اللوجستي، وk-means clustering. وكما هو موضَّح في الشكل ٢-٧؛ فإن للتحليلات الوصفية فرعين رئيسـين هما: الإحصـاءات، ومعالجة التحليلات عبر الإنترنـت (OLAP). يُعَدُّ OLAP هو المصطلح المسـتخدَم لتحليل وتوصيف وتلخيص البيانات المهيكلة المخزَّنة في قواعد البيانات التنظيميـة (غالباً ما يتم تخزينها في مسـتودع البيانات أو في سـوق البيانات، وسـيتمُّ تغطية تفاصيل مسـتودعات البيانات في الفصل ٣) باسـتخدام مكعبات (أيْ: هيـاكل بيانات مُتعدِّدة الأبعاد والتي يتم إنشاؤها لاستخراج مجموعة فرعية من قِيم البيانات للإجابة عن سؤال أعمال معـين). كما أطلـق على فرع OLAP للتحليل الوصفي اسـم ذكاء الأعمال. من ناحيةٍ أخرى؛ تسـاعد الإحصائيات في توصيف البيانات إمَّا على متغـيرٍ واحد في كلِّ مرة أو متغيرات متعددة معًا، باستخدام إما أساليب وصفية أو استنتاجية.



شكل ٢-٧: العلاقة بين الإحصاء والتحليلات الوصفية

كانت الإحصائيات - وهي مجموعة من التقنيات الرياضية لتوصيف وتفسير البيانات - موجودةً منذ وقتٍ طويل. وقد تمَّ تطوير العديد من الأساليب والتقنيات لتلبية احتياجات المستخدمين النهائيين ومعالجة الخصائص الفريدة للبيانات التي يتم تحليلها. وبشكلٍ عام، وعلى أعلى مستوى، فيمكن تصنيف الأساليب الإحصائية على أنها إما وصفية أو استنتاجية. ويكمُنُ الفرق الرئيس بين الإحصاء الوصفي والاستنتاجي في البيانات المُستَخدَمة في هذه الطرق - في حين يقوم الإحصاء الوصفي على وصف بيانات العينة المتاحة، تتعلق الإحصاءات الاستقصائية برسم استدلالات أو استنتاجات حول خصائص السكان. وسنقوم إيجازًا في هذا القسم، بوصف الإحصائيات الوصفية (لأنها بالفعل تضع حجر الأساس للتحليلات الوصفية، وتُعَد كذلك الجزء الأساسي منها)، أما في القسم التالي؛ فسنغطي الانحدار (بنوعيه الخطي واللوجستي) كجزءٍ من الإحصاءات الاستدلالية.

الإحصاء الوصفي للتحليلات الوصفية:

تُوضِّح الإحصائية الوصفية، كما يُوحي الاسم، الخصائص الأساسية للبيانات المتحصَّل عليها، وغالبًا ما يكون هناك متغيرٌ واحدٌ في كل مرة. وباستخدام المعادلات والمجموعات العددية، تستطيع الإحصاءات الوصفية تلخيص البيانات بطريقة تؤدي في الغالب إلى الخروج من الدراسة بأنماط سهلة الفهم وذات معنى. وعلى الرغم من أن الإحصاء الوصفي مفيدٌ جدًا في تحليلات

البيانــات، وأيضًا له الرَّواج الأعلى في الاســتخدام بين الأســاليب الإحصائية؛ فإنه لا يســمح بعمل اســتنتاجات (أو اســتدلالات) تتجاوز عينــة البيانات التي يتــمُّ تحليلها. أيْ: إنه ببســاطة يُعَدُّ طريقــةً جيدةً لتوصيف ووصف البيانات المتاحة، دون الحصول على اســتنتاجات (الاســتدلالات أو الاســتقراءات)؛ وذلك فيما يتعلق بســكان الفرضيات ذات الصِّلة التــي قد نضعها في الاعتبار. وفي تحليــلات الأعــمال؛ تلعب الإحصائيات الوصفية دورًا حساسًــا - فهي تتيــح لنا فهم وشرح/ تقديم بياناتنا بطريقة هادفة باســتخدام الأرقام المجمَّعة أو جداول البيانات أو الخرائط/ الرسوم البيانية. وفي جوهر الإحصائيات الوصفية؛ فهي تساعدنا على تحويل الأرقام والرموز التي بأيدينا إلى تمثيلات ذات معنى يســتطيع أيُّ شــخص أن يفهمها ويســتخدمها بســهولة. لا يساعد هذا الفهم مستخدمي الأعمال في عمليات صُنع قراراتهم فحسب؛ بل يساعد كذلك محترفي التحليلات وعلــماء البيانات عــلى توصيف واختبار صلاحية البيانات لمهامٍّ تحليليــة أخرى أكثر تعقيدًا. كما يســمح الإحصاء الوصفي للمحللين بتحديد التنسيق data concertation، وكذلك القيم الكبيرة أو الصغيرة غير المعتادة (أيْ: القيم المتطرفة)، وقيم البيانات الموزَّعة بشكلٍ غير متوقع للمتغيرات الرقميــة. ولذلك؛ فيمكن تصنيف الطــرق في الإحصاء الوصفي إما كمقاييــس للميل المركزي أو مقاييس للتشــتت. وسنســتخدم في القســم التالي وصفًا بســيطًا وصياغة رياضية/ تمثيلًا رياضيًّا لهذه المقاييس. في التمثيل الرياضي؛ سنســتخدم س١، س٢، س٣، ...، ســن لتمثيل القيم المنفردة (الملاحظات) للمتغير (قياس) الذي نحن بصدد توصيفه.

مقاييس النزعة المركزية (كما يمكن أن يُطلق عليها مقاييس الموقع، أو المقاييس المركزية):

تُعَدُّ المقاييس المركزية؛ هي الطرق الرياضية التي نستخدمها في تقدير ووصف الوضع المركزي للمتغــير المعطى محل الدراســة. ويُعَــدُّ مقياس الاتجاه المركزي قيمةً عدديــةً مفردةً تهدف إلى وصــف مجموعة من البيانات؛ من خلال تحديد أو تقدير الموقع المركزي ضمن البيانات بوضوح. ويُعَدُّ المتوسط (والذي يُطلق عليه غالبًا الوسط الحسابي أو المتوسط البسيط)، هو المقياس الأكثر اســتخدامًا للاتجاه المركزي. بالإضافة إلى المتوســط، من الممكن أيضًا استخدام الوسيط أو المنوال المُستَخدَم لوصف المركزية للمتغير المعطى. وعلى الرغم من أن الوسط الحسابي والوسيط والمنوال كلها مقاييس صحيحة للاتجاه المركزي؛ فإنه في بعض الحالات ووفقًا لظروف معينة، تكون إحدى هــذه المقاييس أكثر ملاءمةً من الأخرى. وفيما يــلي أوصاف مختصرة لهذه المقاييس، بما في ذلك كيفية حسابها رياضيًّا مع الإشارة إلى الظروف التي تجعل أحدها أنسب في الاستخدام من غيره.

- **الوسط الحسابي:** (أو المتوسط البسيط أو المعدل) هو مجموع كلِّ القيم/ الملاحظات مقسومًا على عدد تلك الملاحظات في حزمة البيانات. وهو إلى حَدٍّ بعيد يُعَدُّ الأكثر رواجًا وشيوعًا في قياس النزعة

المركزية. ويتمُّ استخدامه مع بيانات رقمية مستمرة أو منفصلة. بالنسبة للمتغير س، إذا حصلنا على عدد ن من القيم/ الملاحظات (س١، س٢، س٣، ...، س ن)؛ فيمكننا كتابة الوسط الحسابي لعينة البيانات (س، وتنطق س - بار) كما يلي: (يتم استبدال كل (x) بـ س، وكل (n) بـ ن)

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

or

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

يمتلك الوسط العديد من الخصائص الفريدة؛ فعلى سبيل المثال: يكون مجموع الانحرافات المطلقة (الفروق بين الوسط الحسابي والقيم) فوق الوسط هي نفس مجموع الانحرافات التي تحت الوسط، مع مراعاة توازن القيم في كلا الجانبين. ومع ذلك، فإن هذا لا يعني أن نصف عدد القيم يكون أعلى الوسط الحسابي، والنصف الآخر أقل منه (فهذا اعتقادٌ خطأ شائعٌ بين أولئك الذين لا يعرفون الإحصائيات الأساسية). وأيضًا يُعَدُّ الوسط مقياسًا فريدًا لكل مجموعة بيانات، وله معنًى وقابلٌ للحساب لكل ما يتعلق بالبيانات والفترات العددية. ويُعَدُّ أحد أكثر العيوب الرئيسة في الوسط الحسابي؛ هو تأثره بالقيم المتطرفة (القيم التي تكون أكبر أو أصغر بكثيرٍ من بقية نقاط البيانات). إذ تقوم القيم المتطرفة بسحب الوسط نحوها، وبالتالي يحدث تحيُّز في التمثيل المركزي. ولذلك؛ في حالة وجود قيم متطرفة أو إذا كانت البيانات مبعثرة ومنحرفة بطريقة عشوائية؛ فيجب على المحلل تجنُّب استخدام الوسط الحسابي كمقياس مركزي أو إضافته مع مقاييس النزعة المركزية الأخرى، مثل: الوسيط والمنوال.

- **الوسيط:** هو مقياس لقيمة المركز في حزمة البيانات المعطاة. أيْ: هو الرقم الواقع في وسط مجموعة معينة من البيانات التي تمَّ ترتيبها / فرزها حسب حجمها (إما تصاعديًّا أو تنازليًّا). إذا كان عددُ القيم رقمًا فرديًّا؛ فإن تحديد الوسيط سهل للغاية - فقط قم بفرز القيم بناءً على قيمها واختر القيمة الصحيحة التي في الوسط تمامًا. أما إذا كان عدد القيم رقمًا زوجيًّا؛ فقم بتحديد القيمتين الواقعتين في المنتصف، ثم خذ المتوسط البسيط لهاتين القيمتين. يُعَدُّ الوسيط ملائماً وقابلاً للحساب لأنواع معينة من البيانات كالنِّسب، والفترات الفاصلة، والبيانات الترتيبية. وبمجرد تحديد الوسيط، يكون نصف عدد البيانات في الأعلى منه والنصف الآخر أدناه. وعلى عكس الوسط الحسابي، لا يتأثر الوسيط بالقيم المتطرفة أو البيانات المنحرفة.
- **المنوال:** هو القيمة التي تحدث بشكلٍ متكرر (القيمة الأكثر تكرارًا في مجموعة البيانات). ويتمُّ تمثيله على الرسم البياني بالعمود الأعلى، وبالتالي، يمكن اعتباره الخيار/ القيمة الأكثر شيوعًا.

ويكون المنوال مفيدًا جدًا في حالة مجموعات البيانات التي تحتوي على عددٍ صغيرٍ نسبيًا من القيم الفريدة. أي إنه قد يكون عديم الفائدة إذا كانت البيانات تحتوي على الكثير من القيم الفريدة (كما هو الحال في العديد من القياسات الهندسية ذات الدقة العالية، والعدد الكبير من الخانات العشرية)؛ مما يجعل كلَّ قيمة تحتوي على رقمٍ واحد تمثل تكرارًا خاصًّا بها. وعلى الرغم من كونه مقياسًا مفيدًا (خاصة للبيانات الاسمية)؛ فإنه لا يمثل المركزية بصورة جيدة؛ ولذلك لا ينبغي استخدامه كمقياس وحيد للنَّزعة المركزية لمجموعة بيانات معينة.

باختصار؛ أيٌّ من مقاييس الاتجاه المركزية يُعَدُّ الأفضل؟ في الواقع لا تُوجَد هناك إجابةٌ واضحةٌ لهذا السؤال؛ ولكن هناك بعض الإشارات نعرض لها فيما يلي: استخدم الوسط الحسابي عندما لا تكون البيانات ميَّالة للقيم المتطرفة، وليس بها انحرافات معنوية؛ استخدم الوسيط عندما تكون بالبيانات قيم متطرفة و/ أو أن تكون بطبيعتها بيانات ترتيبية؛ استخدم المنوال عندما تكون البيانات اسمية. ولعل أفضل هذه الممارسات هو استخدام كلِّ هذه المقاييس الثلاثة معًا؛ بحيث يمكن الوصول للاتجاه المركزي لمجموعة البيانات، وتمثيله من وجهة نظر كلِّ مقياس على حِدَة. ولأن المعدَّل "أو المتوسط" هو يكون غالبًا مفهومًا مألوفًا للغاية ومُستَخدَمًا بشكلٍ كبيرٍ للجميع في الأنشطة اليومية المتكرِّرة بانتظام؛ فغالبًا ما يستخدم المديرون (وكذلك بعض العلماء والصحفيين) المقاييس المركزية (خاصة الوسط الحسابي) بشكلٍ غير ملائم خاصةً عندما يستلزم الأمر التعامل مع المعلومات الإحصائية الأخرى من منظور المركزية. ومن الأفضل تقديم إحصائيات وصفية كحزمة - عبارة عن مزيجٍ من مقاييس مركزية وتشتُّت - في مقابل مقياسٍ واحدٍ مثل الوسط الحسابي.

مقاييس التشتُّت (وتُسمَّى أيضًا مقاييسُ الانتشار أو الامتداد):

تُعَدُّ مقاييس التشتُّت هي الطرق الرياضية المُستَخدَمة لتقدير أو وَصْف درجة التباين في المتغير المعطى الذي يكون محل الدراسة. وهي تمثل الانتشار الرقمي (الاكتناز أو عدمه) لمجموعة البيانات المعطاة. ولوصف هذا التشتت، تمَّ تطوير عددٍ من المقاييس الإحصائية؛ وأبرزها: النطاق، والتباين، والانحراف المعياري (وكذلك الربيعات، والانحراف المطلق). يُعَدُّ أحد الأسباب الرئيسية وراء أهمية مقاييس تشتُّت/ انتشار قيم البيانات، هو أنه يمنحنا إطارًا يمكننا من خلاله الحكم على النزعة المركزية - وذلك بأن يعطينا إشارةً إلى مدى جودة الوسط الحسابي (أو المقاييس المركزية الأخرى) عند تمثيل بيانات العينة. وإذا كان تشتُّت القيم في مجموعة البيانات كبيرًا؛ فلا يُعتَبر الوسط الحسابي تمثيلًا جيدًا للبيانات. وذلك لأن قياس التشتُّت الكبير يشير إلى اختلافات كبيرة بين الدرجات المنفردة.

وفي البحث أيضًا، غالبًا ما يُنظَر إليه على أنه علامة إيجابية لرؤية تباين صغير داخل كلِّ عينة بيانات؛ إذ إنه قد يشير إلى التجانس، والتشابه، والمتانة ضمن البيانات التي تمَّ جمعها.

- **النطاق:** قد يكون النطاق أبسط مقياس للتشتت. والنطاق هو الفرق بين أكبر قيمة وأصغر قيمة في مجموعة البيانات المعطاة (أيْ: المتغيرات). أي إننا نحسب النطاق ببساطة عن طريق تحديد أصغر قيمة في مجموعة البيانات (الحد الأدنى)، وتحديد أكبر قيمة في مجموعة البيانات (الحد الأقصى)، وحساب الفرق بينهما أيْ: (النطاق = الحد الأقصى - الحد الأدنى).
- **التباين:** هو مقياسٌ أكثر شمولًا وتطوُّرًا للتشتُّت. ويُستخدَم التباين لحساب انحراف جميع نقاط البيانات في مجموعة البيانات المعطاة عن الوسط الحسابي. وكلما كان التباين أكبر، زاد انتشار (تشتُّت) البيانات عن الوسط وأيضًا زاد تنوُّعها بالشكل الذي يسمح لأيِّ شخص ملاحظته في عينة البيانات. ولمنع اختلاط الفروق السلبية والإيجابية، يأخذ التباين مربع المسافات من الوسط الحسابي في الاعتبار. ويمكن كتابة معادلة عينة البيانات كالتالي: (يتم استبدال كلِّ (x) بـ س، وكل (n) بـ ن)

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}$$

حيث ن هي حجم العينة، وس (عليها شرطة) هو الوسط الحسابي للعينة، وس (بجوارها i صغيرة) هو القيمة ذات الترتيب (i) في العينة. تشير قيمُ التباين الأكبر إلى مزيدٍ من التشتُّت؛ في حين تشير القيم الأصغر إلى انضغاط نقاط البيانات في حزمة البيانات الكلية. ونظرًا لأن الفروق مربعة؛ فإن الانحرافات الأكبر عن الوسط الحسابي تُسهِم بشكلٍ معنوي في قيمة التباين. مرةً أخرى، ونظرًا لأن الفروق مربعة؛ فإن الأرقام التي تمثل الانحراف/ التباين تصبح غير ذات معنى إلى حَدٍّ ما (في مقابل الفرق بالدولار، وهنا يتم الحصول على فرق مربع للدولار). لذلك؛ ففي العديد من تطبيقات الأعمال لا نستخدم التباين كمقياس للتشتُّت، ونستعيض عنه بمقياس آخر يعطي للتشتُّت معنًى أفضل، ويُسمَّى الانحراف المعياري.

- **الانحراف المعياري:** هو أيضًا مقياسٌ لانتشار القيم ضِمن مجموعة من البيانات. ويتم حسابه ببساطة عن طريق أخذ الجذر التربيعي للتغيرات. وتوضِّح الصيغة التالية حساب الانحراف المعياري من عينة معينة من نقاط البيانات: (يتم استبدال كلِّ (x) بـ س، وكل (n) بـ ن)

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}}$$

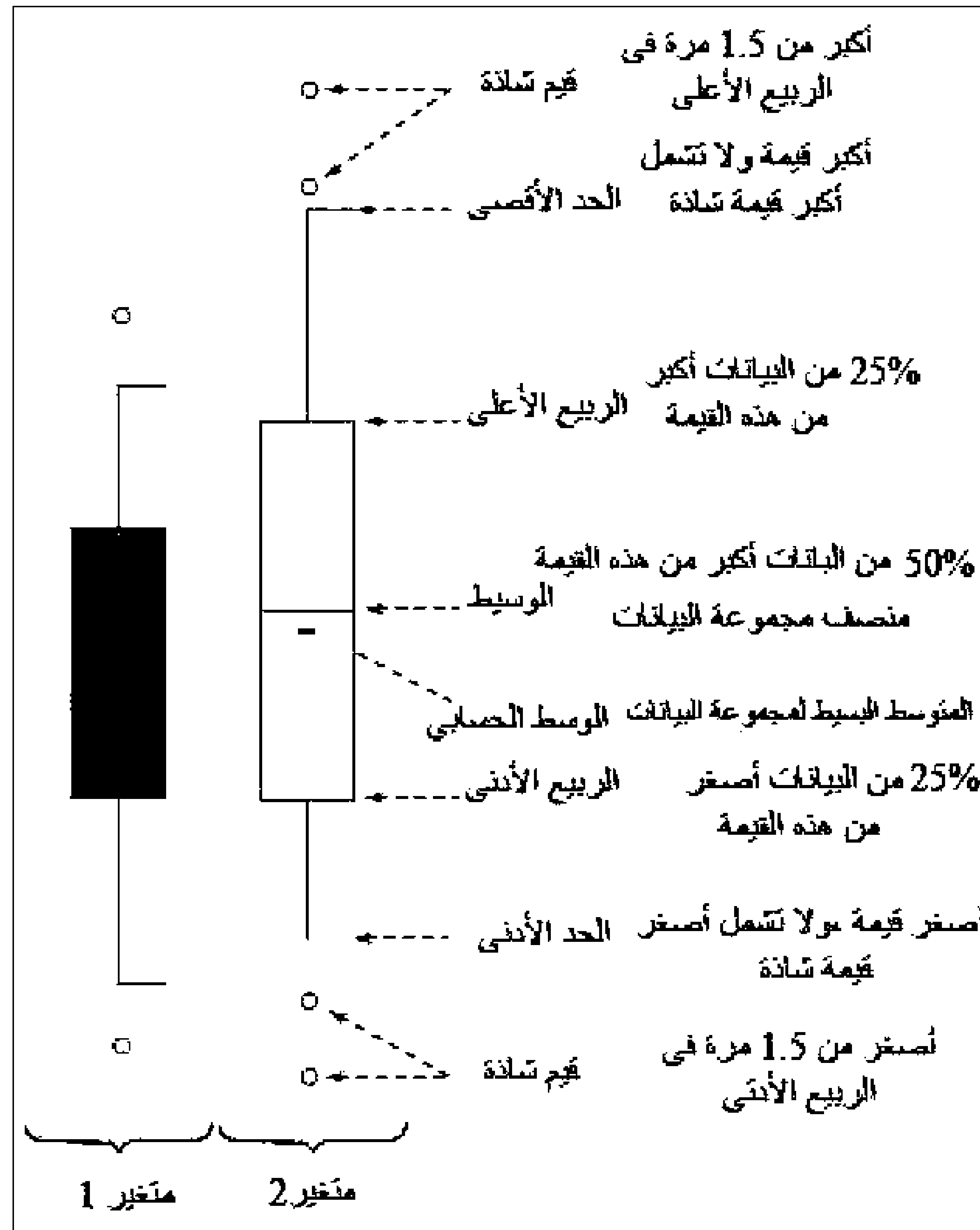
- **متوســـط الانحـــراف المطلق:** إضافةً إلى التبايـــن والانحراف المعياري، نســـتخدم أيضًا في بعض الأحيان متوســـط الانحراف المطلق لقياس التشـــتُّت في مجموعة البيانات. ويُعَدُّ استخدام هذا المقياس طريقة أبســـط لحســـاب الانحراف الإجمالي عن الوسط الحسابي. وعلى وجه التحديد، يتم حســـاب متوســـط الانحراف المطلق من خلال قياس القيم المطلقـــة للفروق بين كل نقطة من نقاط البيانات والوســـط الحســـابي، ثم جمع تلك القيم. ويوفِّر ذلك مقياسًا للانتشار دون أن يكون محددًا حول نقطة بيانات تكون أدنى أو أعلى من الوســـط الحســـابي. وتوضِّح الصيغة التالية حساب متوسط الانحراف المطلق:

$$MAD = \frac{\sum_{i=1}^{n}|x_i - \overline{x}|}{n}$$

- **الرُّبيعـــات والنطـــاق الرُّبيعي:** تســـاعدنا الرُّبيعات على تحديد الانتشـــار داخل مجموعة فرعية من البيانات. ويشـــكِّل الرُّبيع ربعَ عدد نقاط البيانـــات المعطاة في مجموعة البيانات. ولتحديد الرُّبيعـــات يتـــمُّ أولًا ترتيب البيانات، ثـــم بعد ذلك يتم تقســـيم تلك البيانـــات المرتبة إلى أربع مجموعـــات بيانات أصغر منفصلة بعضها عن بعض. وتُعَدُّ الرُّبيعات مقياسًـــا مفيدًا للتشـــتُّت؛ لأنها - مقارنة بالمقاييـــس المناظرة الأخرى - أقل تأثرًا بالقيم المتطرفة أو الانحرافات في مجموعة البيانات بالكامل. وغالبًا ما يتمُّ اســـتخدام الرُّبيعات كأفضل خيار لقياس التشتُّت كما يتم اعتبار الوسيط هو مقياس التشتُّت الأنسب لقياس الاتجاه المركزي؛ وذلك عند التعامل مع الانحراف و/ أو البيانات ذات القيم المتطرفة. إنَّ الطريقة الشائعة للتعبير عن الرُّبيعات هي النطاق (المدى) الرُّبيعـــي، والتي تصف الفرق بين الربيع الثالـــث (Q3) والربيع الأول (Q1)، وهذا الفرق يعطي دلالةً على مدى النصف الأوسط من الدرجات في التوزيع. ويُعَدُّ أفضل شرح أو توضيح للمقاييس الوصفية المســـتندة على الرُّبيعات (ســـواء كانت مقاييس مركزية أو مقاييس تشتت)، هو وجود ما يُسمَّى popular plot، والذي يُطلَق عليه أيضًا a box plot أو (box-and-whiskers plot).

- **مُخطط الصندوق وطرفاه Box-and-Whiskers Plot أو ببســـاطة الرســـم الصندوقي a box plot:** هو رســـمٌ بيـــاني يُوضِّح العديد من الإحصائيـــات الوصفية حول مجموعة البيانـــات المعطاة. ومن الممكن أن يكون هذا الرســـم إما أفقيًّا أو رأســـيًّا؛ ولكن الرأسي هو التمثيل الأكثر شيوعًا؛ خاصةً في منتجات برمجيات التحليلات الحديثة. ومن المعروف أنَّ أول مَنْ أنشأ هذا المخطط وقدَّمه هـــو John W. Tukey في عام ١٩٦٩م. ويُوضِّح الشـــكل ٢-٨ زوجـــين من box plots جنبًا إلى

جنب، يتشـاركان نفس المحور Y (أو ص على حسـب الطباعة للشـكل). كما هو موضح هنا، يمكـن أن يحتـوي مخططاً واحداً على box plot واحداً أو أكثر لأغراض المقارنة المرئية. في مثل هذه الحالات؛ يكون المحور y (أو ص على حسب الطباعة للشكل)؛ هو المقياس الشائع للحجم (القيمة العددية للمتغير)، في حين يُعبِّر المحور السـيني (أو x) عن فئات مختلفة/ مجموعات فرعيـة مختلفـة كالأبعاد الزمنية المتفاوتة (مثـل: إحصائيات وصفية لنفقـات الرِّعاية الطبية السـنوية في عام ٢٠١٥م مقابل عام ٢٠١٦م) أو فئات مختلفة (مثل: الإحصاء الوصفي لنفقات التسويق مقابل إجمالي المبيعات). وعلى الرغم من أنه تاريخيًّا، لم يتم استخدام box plot على نطاق واسـع بما يكفي في الغالب (وخاصة في المناطق خارج الإحصاءات)، ومع الرَّواج الحاصل لتحليلات الأعمال الناشـئة؛ فإنه (أيْ: box plot) يكتسب الشـهرة في المجالات الأقل تقنيةً في عالم الأعمال. ويؤدي ثراء المعلومات الناتجة من box plot، وأيضًا سـهولة فهمها إلى حَدٍّ كبيرٍ إلى زيادة الثقة فيه ورواج اسـتخدامه في الآونة الأخيرة. يُوضِّح box plot المركزية (الوسـيط وأحيانًا أيضًا الوسـط الحسابي)، وكذلك التشتُّت (كثافة البيانات في النصف الأوسط، والمرسومة كصنـدوق بين الربعين الأول والثالث)، والحدان الأدنى والأقصى (والموضحة كخطوط ممتدة من الصندوق، تبدو كالشوارب، والتي يتمُّ حسابها على أنها ١,٥ أضعاف الطرف العلوي أو السفلي من الصندوق الرُّبيعي) إلى جانب القيم الشـاذة الأكبر من حدود الشوارب. يوضِّح box plot أيضًا ما إذا كانت البيانات مُوزَّعة بشـكلٍ متماثلٍ فيما يتعلق بالوسـط الحسابي أو أنها تتمايل بطريقة أو بأخرى. ويعطي الموضع النسبي للوسيط مقابل الوسط الحسابي وأطوال الشعيرات الممتدة على جانبي الصندوق دلالةً جيدةً على الانحراف المحتمل في البيانات.

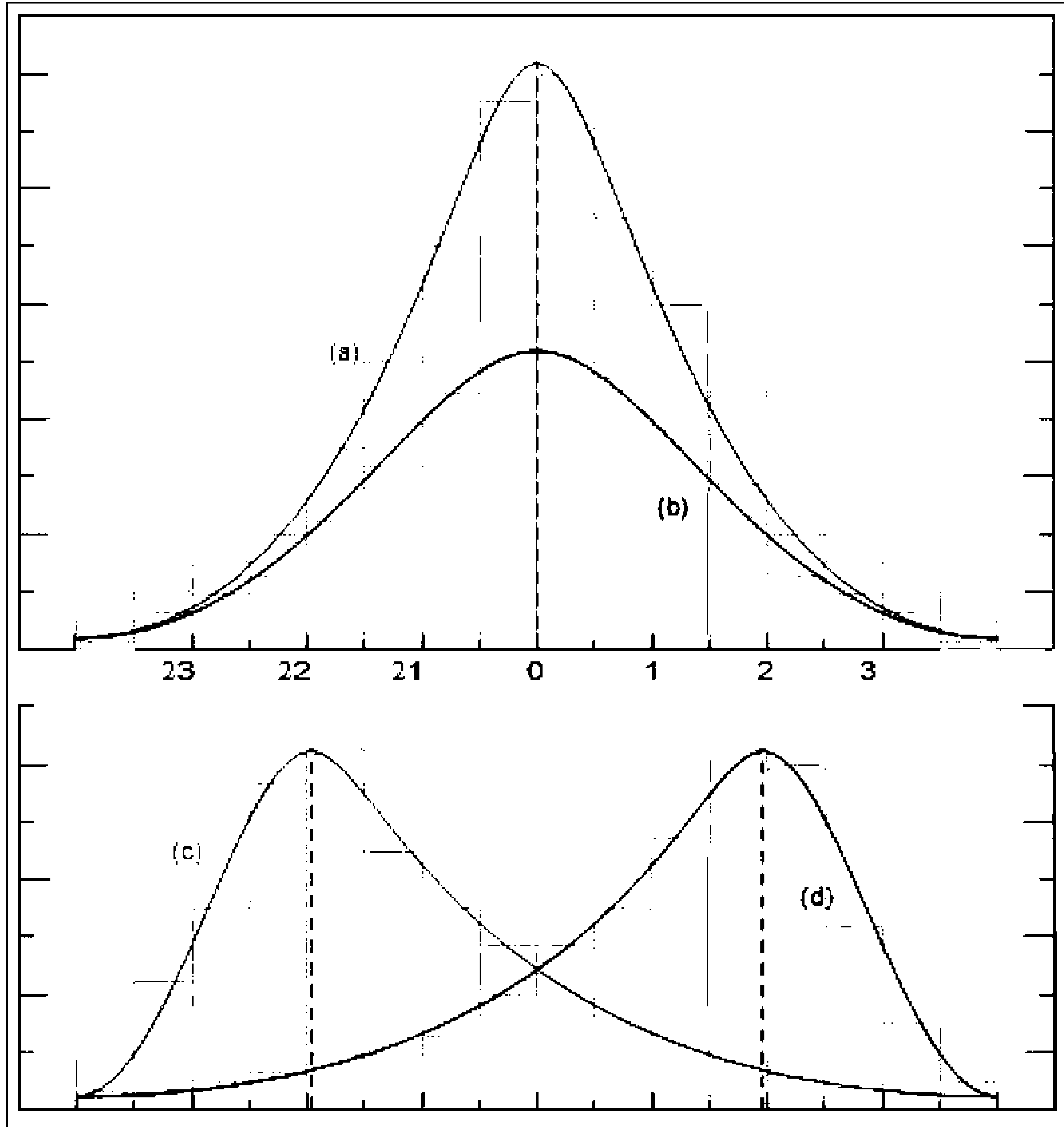


شكل ٢-٨: فهم التفاصيل عن مخططات **Box-and-Whiskers**

شكل التوزيع:

على الرَّغم من أنه ليس شائعًا مثل المركزية والتشتت؛ فإن شكل توزيع البيانات يُعَدُّ أيضًا مقياسًا مفيــدًا للإحصاءات الوصفية. ولكن قبل الخوض في شــكلِ التوزيــع؛ نحتاج أولًا إلى تعريف التوزيع نفسه. ببساطة نقول: إنَّ التوزيع هو تكرار نقاط البيانات التي يتمُّ حسابها ورسمها على عددٍ صغيرٍ من العلامات الفئوية أو النطاقات العددية (أيْ: صناديق التخزين). وعند رسم التوزيع بيانيًّا، يعرض المحور y (أو ص) التكرار (عدد أو نســبة مئوية)، في حين يعرض المحور x (أو س) الطبقات المنفردة أو الصناديق بنمط ترتيبي. يُســمَّى التوزيع المعــروف جيدًا بالتوزيع الطبيعي، والذي يتطابق تمامًا على كلا جانبي الوسط الحسابي، ويحتوي التوزيع الطبيعي على العديد من الخصائص الرياضية التي

تمَّ إنشاؤها بشكلٍ جيد والتي تجعله أداةً مفيدةً جدًا للبحث والممارسة. وكلما زاد تشتُّت مجموعة البيانات؛ زاد الانحرافُ المعياري، وكذلك يبدو شــكل التوزيع أعرض. يعرض الشكل ٢-٩ رسماً بيانياً للعلاقة بين التشتُّت وشكل التوزيع (وذلك في سياق الكلام عن التوزيع الطبيعي).



شكل ٢-٩: العلاقة بين التشتت وخصائص الشكل

يُوجَد نوعان من المقاييس الشائعة الاستخدام لحساب خصائص شكلِ التوزيع وهما: الانحراف والتفرطح. وغالبًا ما يستخدم الرسم البياني (مخطط التكرار) لتوضيح كلٍّ من الانحراف والتفرطح.

- الانحراف: هو قياسٌ لعدم التماثل (التمايل) في توزيع البيانات التي تصوِّر هيكلاً أحاديَّ الشكل - في هــذا التوزيــع للبيانات تُوجَد قمة واحدة فقــط. ولأن التوزيع الطبيعي هو توزيعٌ أحاديٌّ

الشكل متماثلٌ تمامًا؛ فلا يُوجَد به انحرافات؛ أي إنَّ قياس الانحراف (أو قيمة معامل الانحراف) يساوي الصفر. ويمكن أن يكون قياس/ قيمة الانحراف إما بالموجب أو بالسالب. إذا مال التوزيع ناحية اليســار (أي إنَّ الذيل على الجانب الأيمن وكان الوسط الحسابي أصغر من الوسيط)؛ فإنه ينتــج عنه مقياس انحراف موجب، أما إذا مال التوزيع ناحية اليمين (أي إنَّ الذيل على الجانب الأيسر وكان الوسط الحسابي أكبر من الوسيط)؛ فعندئذٍ يكون الانحراف سالباً. ففي الشكل ٢-٩؛ يمثل المنحنى (c) توزيعًا منحرفًا موجبًا، في حين يمثل المنحنى (d) توزيعًا منحرفًا سالبًا. وفي نفس الشكل، يمثل كلٌّ من (a) و(b) تماثلاً مثالياً، وبالتالي قياس الانحراف يساوي صفر.

$$Skewness = S = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{(n - 1)s^3}$$

حيث s هو الانحراف المعياري، وn هي حجم العينة.

- **التفرطح:** هو مقياسٌ آخر لاستخدامه في تحديد خصائص شكل التوزيع الأحادي. وبغض النظر عن ميل شــكل التوزيع؛ فإنَّ التفرطــح أكثر اهتمامًا بتمييز طبيعــة التوزيع من حيث القمة/ الطــول/ النحافة. وبشــكلٍ أكثر تحديدًا؛ فــإنَّ التفرطح يقيس درجة ارتفــاع أو انخفاض قمة التوزيــع عنها في التوزيع الطبيعي. في حين يشــير التفرطح الموجــب إلى توزيع مدبب/ طويل نســبيًّا، ويشير التفرطح السَّلبي إلى توزيع مسطح/ قصير نسبيًا. وكنقطة مرجعية؛ فإن التوزيع الطبيعي له تفرطح ٣. ويمكن كتابة معادلة التفرطح كالتالي:

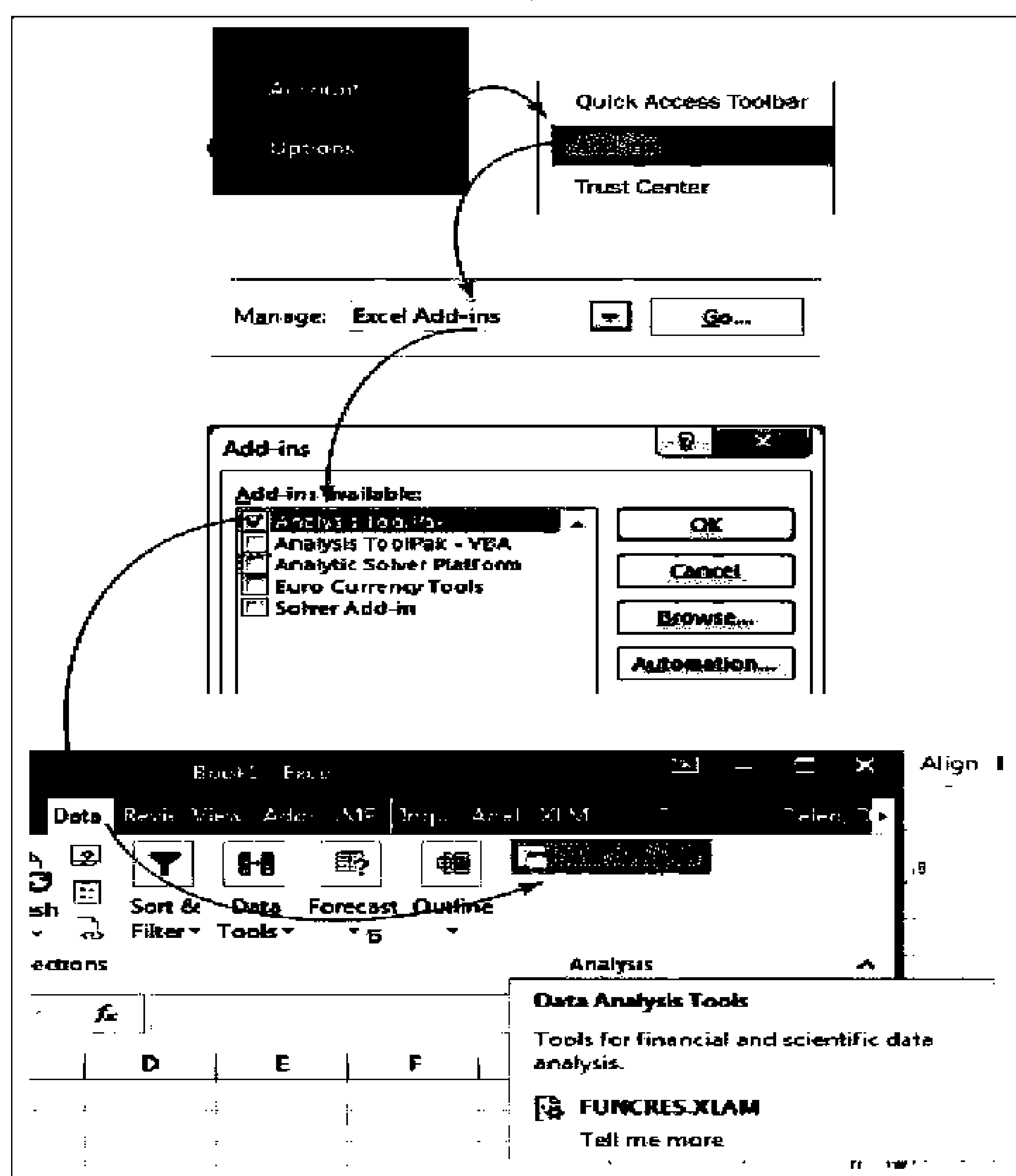
$$Kurtosis = K = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{ns^4} - 3$$

إنَّ الإحصائيات الوصفية (شــأنها شــأن الإحصائيات الاســتنتاجية) يمكن حســابها بسهولة؛ وذلــك باســتخدام حــزم برمجيــة إحصائيــة ذات جــدوى تجاريــة (مثــل: SAS، أو SPSS، أو Minitab، أو JMP، أو Statistica) أو البرامج المجانية/ المفتوحة (مثل: R). ولعل الطريقة الأكثر ملاءمةً لحســاب الإحصائيات الوصفية، وبعض الإحصائيات الاســتنتاجية هو اســتخدام جداول الإكســل. وتوضِّح الفكرة التقنية ٢-١ بالتفصيل كيفية اســتخدام جداول مايكروسوفت إكسل لحساب الإحصائيات الوصفية.

رؤية فنية ٢-١

كيفية حساب الإحصائيات الوصفية باستخدام جداول مايكروسوفت إكسل

يمكِن القول بأن برنامج إكسـل يُعَدُّ أداة تحليل البيانات الأكثر شيوعًا في العالم، كما يمكن اسـتخدامه بسهولةٍ بالنسـبة للإحصاءات الوصفية. وعلى الرغم من أن التكوين الأسـاسي لبرنامج إكسـل لا يُوحي أن به دوال إحصائية متاحة للمسـتخدمين النهائيين؛ فـإنَّ هذه الدوال يمكن تثبيتها وتنشـيطها (عند التشـغيل) بالقليـل فقط من نقرات الماوس. ويوضِّح الشـكل ٢-١٠ كيفية تنشيط هذه الدوال الإحصائية (كجزءٍ من تحليل ToolPak) في مايكروسوفت إكسل ٢٠١٦م.

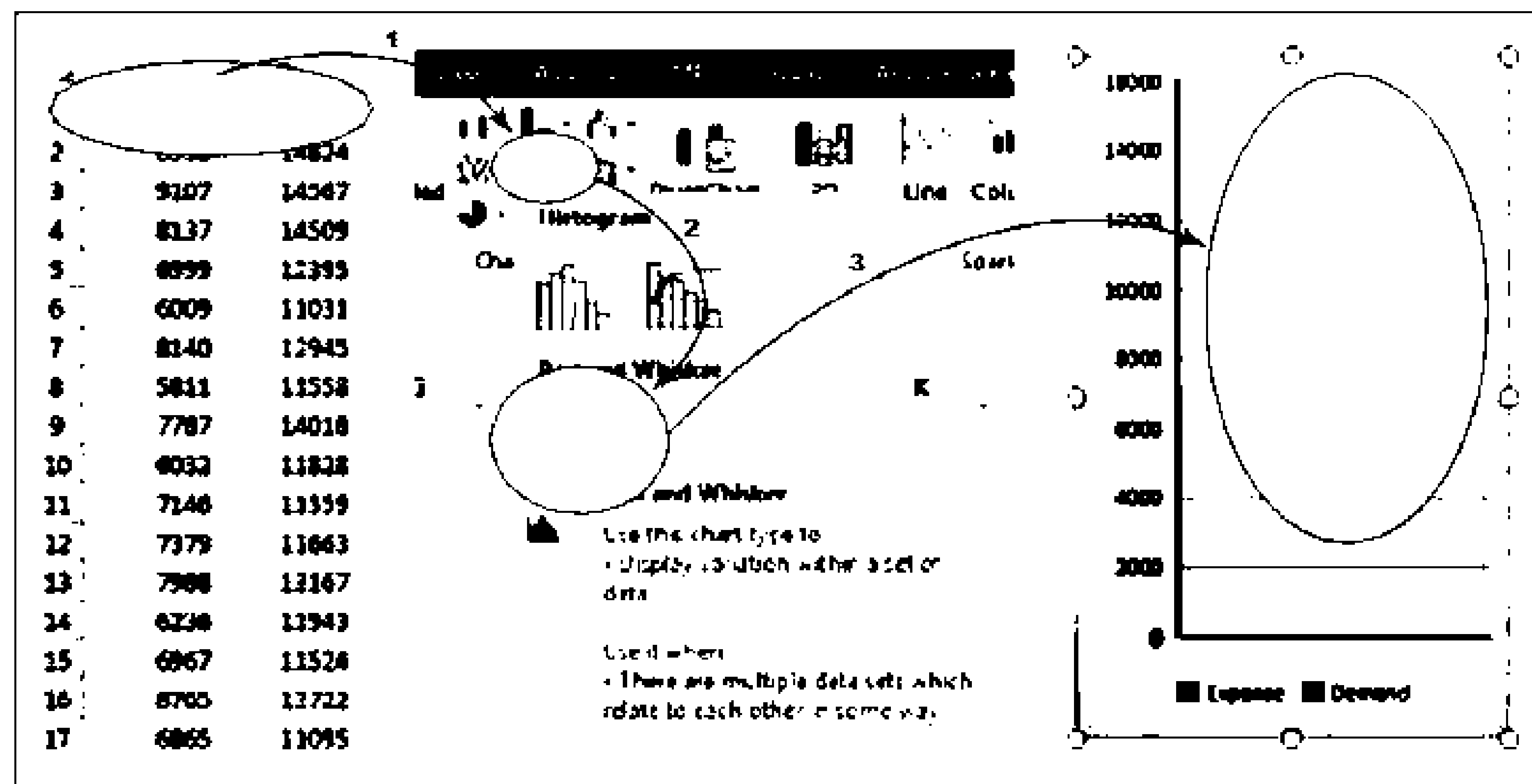


شكل ٢-١٠: تفعيل الدوال الإحصائية في إكسل ٢٠١٦

وبمجرد تفعيل الدوال الإحصائية؛ سيظهر تحليل ToolPak في خيار قائمة البيانات (Data) تحت اسم تحليل البيانات (Data Analysis). وعند النقر فوق "تحليل البيانات أو (Data Analysis)" في مجموعة "التحليل أو (Analysis)" تحت علامة التبويب "البيانات أو (Data)" في شريط قوائم إكسل، ستشاهد "الإحصائيات الوصفية أو (Descriptive Statistics)" كأحد الخيارات الموجودة ضمن قائمة أدوات تحليل البيانات (انظر: الشكل ٢-١١، الخطوات [١ و٢])؛ انقر فوق "موافق أو (OK)"، وسيظهر مربع حوار بعنوان "الإحصائيات الوصفية (Descriptive Statistics)" (انظر: منتصف الشكل ٢-١١). في مربع الحوار هذا؛ تحتاج إلى إدخال نطاق البيانات، والتي يمكن أن يتمَّ تمثيلها بواحدٍ أو أكثر من الأعمدة الرقمية، إلى جانب مربعات فحص التفضيلات، وانقر فوق OK (انظر: الشكل ٢-١١، الخطوات [٣ و٤]). إذا تمَّ تظليل (Selection) أكثر من عمود رقمي واحد؛ فستقوم الأداة بمعالجة كلِّ عمود على أنه مجموعة بيانات منفصلة، ومن ثمَّ ستُقدم إحصائيات وصفية لكل عمود على حِدَة.

شكل ٢-١١: الحصول على الإحصائيات الوصفية في برنامج إكسل

وكمثالٍ بسـيطٍ، اخترنا عمودين (تـمَّ وَضْع عنوانين لهما: مصاريف وطلبات) ونفَّذنا خيار الإحصاء الوصفي (Descriptive Statistics). يُوضِّح الجزء السـفلي من الشـكل ٢-١١ المخرجات التي تمَّ إنشـاؤها بواسطة إكسل. وكما يتضح؛ أنتج إكسـل جميع الإحصائيات الوصفية التي تمَّت تغطيتها في القسم السابق، وإضافة المزيد منها إلى القائمة. وفي إكسل ٢٠١٦م، من السهل جدًا أيضًا (بقليلٍ من نقرات الماوس) أن يتمَّ إنشـاء a box-and-whiskers. كما يُوضِّح الشكل ٢-١٢ العملية البسيطة والمكونة من ثلاث خطوات لإنشاء abox-and-whiskers في إكسل.



شكل ٢-١٢: إنشاء مخطط Box-and-Whiskers في إكسل ٢٠١٦

وعـلى الرغم مـن أنَّ هذه الأداة تُعَدُّ مفيدةً جدًا في إكسـل؛ فإنه يتحتَّم على المحلِّل أن يكون ملمًّا بنقطة مهمة تتعلق بالنتائج المتحصَّل عليها بواسطة تحليل ToolPak، والتـي تنتهج سـلوكًا مختلفًا عن وظائف الإكسـل الأخـرى المعتادة. وعلى الرغم من أن وظائف إكسـل تتغيَّر ديناميكيًّا مثلما تتغيَّر البيانات الأساسية في جدول البيانـات؛ فإنه لا يتم تغيير النتائج التي تم إنشـاؤها بواسـطة تحليل ToolPak. وعلى سـبيل المثال: إذا قمتَ بتغيير القيم في أيٍّ من هذين العمودين أو كليهما؛ فسـتبقى نتائـج الإحصائيات الوصفية التي أنتجهـا تحليل ToolPak كما هي دون تغيير. ومع ذلك؛ لا ينطبق الأمر نفسه على وظائف إكسل المعتادة. إذا كنتَ قد قمتَ بحسـاب قيمة الوسـط الحسـابي لعمودٍ ما (باستخدام الدالة AVEREGE) للنطـاق مـن (A1:A121)، ثـم قمتَ بتغيير القيـم داخل نطاق البيانات؛ فسـوف تتغير قيمة الوسط الحسابي تلقائيًّا. أي إنه باختصار، لا تحتوي

النتائج التي ينتجها تحليل ToolPak على رابط ديناميكي للبيانات الأساسـية، وإذا تم تغيير أيِّ بيان؛ فيجب إعادة التحليل مرةً أخرى باستخدام مربع الحوار.

تغطـي التطبيقـات الناجحة لتحليـلات البيانات نطاقًا واسـعًا من إعدادات الأعمال والتنظيمات، ومعالجة المشكلات التي كان يُعتقد في السابق أنها غير قابلة للحل. وتمثـل الحالة العملية ٢،٣ مثالًا ممتازًا لهـذه القصص الناجحة؛ إذ تتبَنَّى إدارة بلدية صغيرة منهجَ تحليلِ البيانات لكشف وحل المشكلات بذكاء من خلال تحليل أنماط الطلب والاستهلاك بشكلٍ مستمر.

حالة عملية ٢-٣

مدينة كاري تستخدم التحليلات لتحليل البيانات من أجهزة الاستشعار، وتقييم الطلب والكشف عن المشكلات

إنَّ تسرُّب المياه من الحنفيات، والخلل بغسالة الصحون، وطفايات الحريق المكسورة؛ تُعَـدُّ من أكثر المشـكلات المؤرقـة لأيِّ مالك منزل أو ممارس لأعـمال تجارية إلى أن يتمَّ إصلاحهـا. وهذه المشـكلات قد تكون مكلفةً، ولا يمكن التنبؤ بها، وللأسـف، يكون من الصعب تحديدها. وقد تمَّ تسـهيل عملية اكتشـاف وإصلاح المشكلات التي تؤدي إلى نقصـان أو فقدان الماء في مدينة كاري بشـمال كارولينا؛ وذلـك من خلال مجموعة من عـدَّادات المياه اللاسـلكية وبوابة إلكترونية معتمدة على البيانات يسـهل الوصول إليها بواسـطة جميع العملاء. في هذه العملية، حصلت المدينـة على رؤية ذات صورة كبيرة عن حساسـية استخدام المياه؛ وذلك لتخطيط التوسُّـعات في محطات المياه المستقبلية، وتعزيز جهود الصيانة المُستَهدَفة. عندما قامت مدينة كاري بتثبيت العدَّادات اللاسلكية لعدد ٦٠٠٠٠ مشـترك في عام ٢٠١٠م؛ علمت إدارة البلدة أنَّ التقنية الجديدة لن تقتصر على توفير المال عن طريق إزالة القراءات الشـهرية يدويًّـا؛ بل أدركت المدينة أيضًا أنها سـتحصل على معلومات أكثر دقةً وفي الوقت المناسب حول استهلاك المياه. يقرأ النظام اللاسـلكي Aquastar العدَّادات مرةً واحدة كلَّ سـاعة؛ مما يعنـي الحصول على ٨٧٦٠ نقطة بيانات لكلِّ عميلٍ كل عام بدلًا من قراءة شهرية واحدة، ما يعني ١٢ نقطة بيانات فقـط لكل عميل على مـدار العام. كان للبيانـات قدرات هائلـة، إذا كان من الممكن الحصول عليها وتحليلها بسـهولة. يقول المدير المالي لإدارة بلدة كاري كارن ميلز: «إن القراءات الشهرية لعدادات المياه تشبه الحصول على جالون من بيانات المياه المهمة، أمَّا

القراءة كل ســاعة للعدادات فتشــبه إلى حَدٍّ كبير الحصول على حمام سباحة من الحجم الأولمبي مملوء بالبيانات. كما يســاعدنا برنامج SAS في إدارة حجم تلك البيانات بشكلٍ رائــع". في الواقع؛ فــإن هذا الحل يُمكِّن إدارة المدينة من تحليل عدد نصف مليار نقطة بيانات تخصُّ استخدام المياه وجعلها متاحةً وسهلةً الاستخدام لجميع العملاء.

إن القراءة المباشرة كلَّ ســاعة للعدَّادات ســواء المنزليــة أو التجارية أنتجت بعض التطبيقات العملية جدًا، وهي:

- يمكن لإدارة البلدة إخطار العملاء بأيِّ تسريبات مُحتمَلة في غضون أيام.

- يمكن للعملاء تلقِّي تنبيهات تخطرهم في غضون ساعات بوجود ارتفاع في استخدام المياه.

- يمكن للعملاء مراقبة معدَّلات استهلاكهم للمياه عبر الإنترنت؛ مما يساعدهم على المبادرة في الحفاظ على المياه.

مــن خلال البوابة الإلكترونية؛ شــهدت إحدى المؤسســات التجارية في مدينة كاري ارتفاعًــا حادًّا في اســتهلاك المياه في عطلات نهاية الأســبوع، في الوقــت الذي يقضي فيه الموظفــون تلك العطلات خارج المدينة. كان ذلــك أمرًا غريبًا ومُلفتًا، ولكن القراءة غير العادية للعدَّادات ســاعدت الشركة على اكتشاف خلل في غسالة الصحون التجارية؛ مما يجعلها تعمل بشــكلٍ مستمر خلال عطلة نهاية الأسبوع. وبدون بيانات عدادات المياه اللاسلكية والبوابة الإلكترونية المتاحة للعملاء، كان من الممكن أن تحدث هذه المشكلة دون أن يلاحظها أحدٌ؛ مما يؤدي إلى استمرار إهدار الماء والمال.

استخدام تلك البوابة الإلكترونية أعطى المدينة رؤيةً أكثر دقةً عن الاستخدام اليومي للمياه لكلِّ شــخص، وهو أمرٌ حســاسٌ في التخطيط للتوسُّــعات المستقبلية في منشآت الميــاه. ولعل الميزة الأكــثر لفتًا للانتباه هي أن المدينة تمكَّنَت مــن التحقُّق من أمرٍ ما كان لــه تأثيرٌ بعيد المدى على تكاليف فروع الشركة المختلفة، ألا وهو أن ســكان كاري اقتصاديــون جدًا في اســتخدامهم للميــاه. وتقول مديرة الموارد المائيــة في المدينة ليلى جودوين: "إننا نقوم بحســاب ذلك باستخدام أجهزة حديثة ذات كفاءة عالية، ويمكن أن يصل استخدام المياه في الأماكن المغلقة إلى ٣٥ جالونًا لكل شخصٍ في اليوم الواحد. في حين يبلغ معدل اســتخدام سكان كاري ٤٥ جالونًا، وهو لا يزال يُعَدُّ معدلًا منخفضًا إلى حدٍّ كبير". لماذا يُعَدُّ هذا مهمًّا؟ كانت المدينة تنفق الأموال لتشــجيع ترشــيد استخدام الميــاه، مثل: تقديم خصومات على المراحيض المنخفضة التدفق أو عمل تخفيضات على

براميل الأمطار. والآن يمكن اتخاذ نهج أكثر اســتهدافًا، ومســاعدة مستهلكين مُحدَّدين على فهم وإدارة استخدام المياه سواء في الأماكن المغلقة أو المفتوحة.

لم يكــن برنامج SAS على قدرٍ من الأهمية لتمكين الســكان من فهم اســتخدامهم للمياه فحسب؛ ولكنه أيضًا لعب دورًا خفيًّا لربط قاعدتي بيانات متباينتين. يقول ميلز: "لدينا قاعدة بيانات للفواتير وقاعدة بيانات أخرى لقراءة العدادات. لقد كنا في حاجة إلى تطويرهما معًا وجعلهما قابلتين للتطبيق".

وتُقدِّر إدارة المدينة أنه بمجرد إزالة الحاجة للقراءات اليدوية؛ سيوفر نظام Aquastar أكــثر من ١٠ مليون دولار فوق تكلفــة المشروع. ولكن يمكن أن يوفــر مكوِّن التحليلات وفورات أكبر. وبالفعل؛ فقد استطاع كلٌّ من المدينة والمواطنين الأفراد توفير قدرٍ من المال عن طريق الاكتشــاف المبكر لتسرُّب المياه. وبينما تمضي قُدُمًا في تخطيط احتياجات البنية التحتية المســتقبلية؛ فإن حصول مدينة كاري على معلومات دقيقة حول اســتخدام المياه سيساعدها على الاستثمار في العدد الصحيح من البنية التحتية في الوقت المناسب. وإضافةً إلى ذلك؛ فإن فهم استخدام المياه سيساعد المدينة إذا لحق بها شيءٌ ضار مثل الجفاف.

تقــول ليلى جودوين: "لقــد مررنا بالجفاف في عــام ٢٠٠٧م، وإذا حدث هذا مرةً أخرى، فلدينا خطة لاســتخدام بيانات Aquastar؛ لمعرفة كمية المياه التي نســتخدمها على أســاس يومي والتواصل مــع العملاء. يمكننا عرضُ ما يحــدث هنا، وإليك عزيزي المواطــن المقــدار الذي يمكنك اســتخدامه حال انخفاض منســوب المــاء لدينا. ونأمل ألا نضطر أبدًا إلى استخدام ذلك؛ لكن علينا أن نكون على استعداد دائم».

أسئلة للمناقشة:

١- ما التحديات التي كانت تواجهها مدينة كاري؟

٢- ما الحل الذي تمَّ اقتراحه؟

٣- ماذا كانت النتائج؟

٤- ما المشكلات الأخرى وحلول تحليل البيانات التي تتوقعها لمدن مثل كاري؟

Source: "Municipality puts wireless water meter-reading data to work (SAS® Analytics) - The Town of Cary, North Carolina uses SAS Analytics to analyze data from wireless water meters, assess demand, detect problems and engage customers."

أسئلة مراجعة على القسم ٢-٥:

١- ما العلاقة بين الإحصاءات وتحليلات الأعمال؟

٢- ما الفروق الأساسية بين الإحصاء الوصفي والاستنتاجي؟

٣- اذكر وعرِّف باختصار مقاييس النزعة المركزية للإحصاءات الوصفية.

٤- اذكر وعرِّف باختصار مقاييس التشتُّت للإحصاءات الوصفية.

٥- ما هو مخطط الصندوق وطرفيه؟ وما هي أنواع المعلومات الإحصائية التي يمثلها؟

٦- ما هي خصائص الشكل الأكثر شيوعًا لوصف توزيع البيانات؟

٢-٦ نمذجة الانحدار للإحصاءات الاستدلالية:

ربما يكون الانحدار - وخاصةً الانحدار الخطي - هو أوسع تقنيات التحليلات شهرةً وأكثرها استخدامًا في الإحصاءات. وتعود جذور الانحدار تاريخيًّا إلى عشرينيات وثلاثينيات القرن العشرين، والتي بدأت بالعمل المبكر على الخصائص الموروثة للبازلاء الحلوة من قبل السِّير فرانسيس جالتون ومن بعده كارل بيرسون. ومنذ ذلك الحين؛ أصبح الانحدار هو التقنية الإحصائية لتوصيف العلاقات بين المتغير/ المتغيرات التفسيرية (مدخلات) ومتغير/ متغيرات الاستجابة (مُخرجات).

كما هو الحال في الأساس؛ فإن الانحدار هو تقنية إحصائية بسيطة نسبيًّا لنمذجة تبعية متغير (الاستجابة أو المخرجات) إلى متغيّر تفسيري واحد أو أكثر (مدخلات). وبمجرد تحديدها؛ يمكن تمثيل هذه العلاقة بين المتغيرات بشكلٍ رسمي على أنها دالة خطية/ مضافة أو على أنها معادلة خطية/ مضافة. وكما هو الحال مع العديد من تقنيات النمذجة الأخرى؛ يهدف الانحدارُ إلى التعرُّف على العلاقة الدالية فيما بين خصائص العالم الحقيقي ووصف هذه العلاقة بنموذج رياضي، والذي قد يُستَخدَم بعد ذلك لكشف وفهم تعقيدات الواقع واستكشاف وشرح العلاقات أو التنبؤ بالحوادث المستقبلية.

ويمكن استخدام الانحدار لواحدٍ من غرضين: الأول هو اختبار فرضية أو استقصاء العلاقات المحتملة بين المتغيرات المختلفة، والثاني تقدير التنبؤ بقيم متغيرات الاستجابة بناءً على متغير تفسيري واحد أو أكثر. ولا يقوم أحد هذين الاستخدامين مقام الآخر؛ إذ تُعَدُّ القوة التفسيرية للانحدار هي أيضًا الأساس لقدرتها على التنبؤ. ففي اختبار الفرضية (بناء النظرية)؛ يمكن أن يكشف تحليل الانحدار عن وجود/ قوة واتجاهات العلاقات بين عددٍ من المتغيِّرات التفسيرية (ويُرمَز لها عادةً بالرمز x «أو س») ومتغير الاستجابة (ويرمز له عادةً بالرمز y «أو ص»). أما

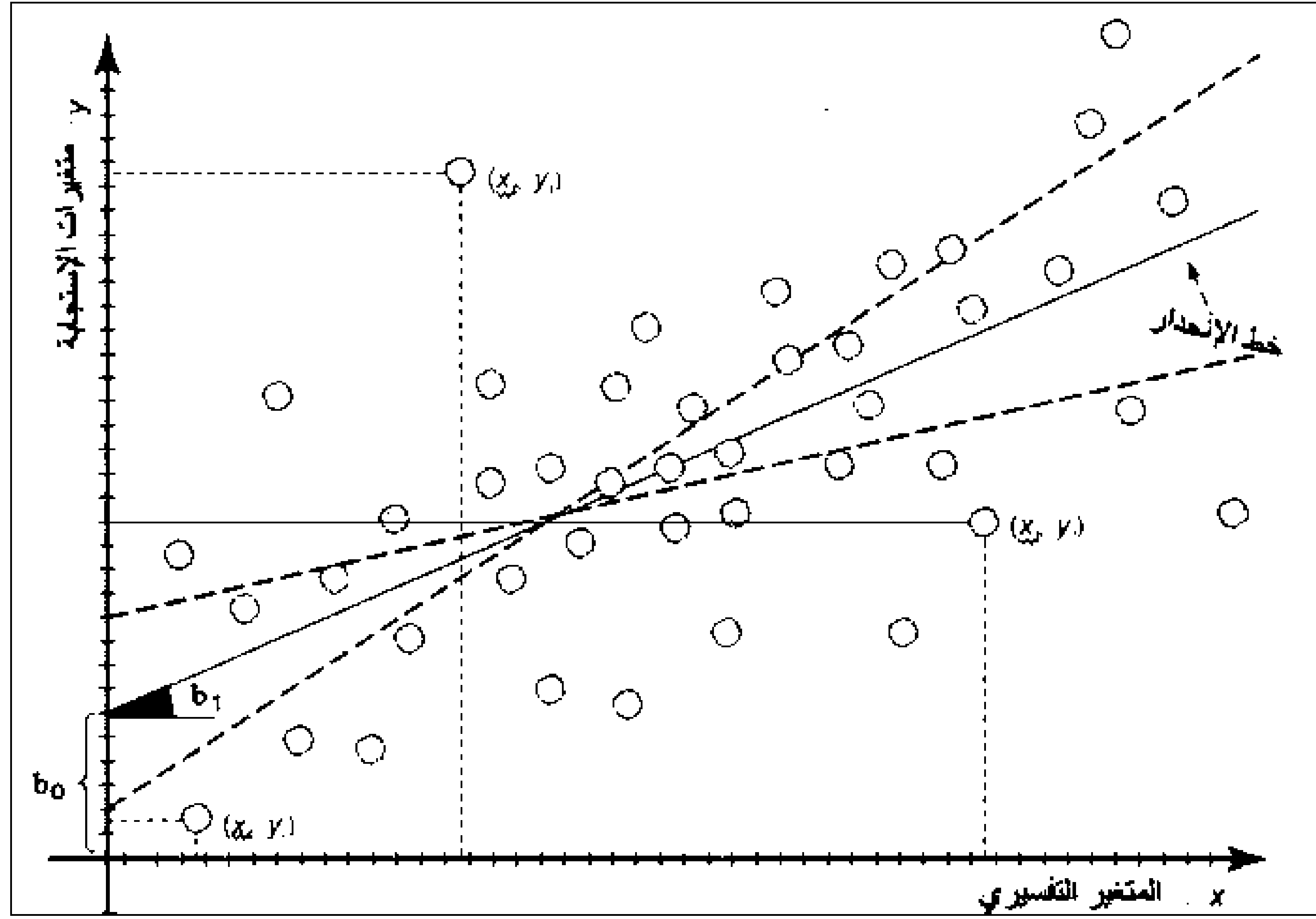
في التنبـؤ؛ فيُحدِّد الانحدار العلاقات الرياضية المضافة (في شـكل معادلـة) بين واحدٍ أو أكثر من المتغيرات التفسيرية ومتغير الاستجابة. وبمجرد تحديد هذه المعادلة؛ يكون من الممكن استخدامها للتنبؤ بقيم متغير الاستجابة لمجموعة معينة من قِيم المتغيرات التفسيرية.

- **الارتبــاط مقابل الانحدار:** ولأن تحليل الانحدار نشـأ من دراسـات الارتبـاط، ولأن كلتا الطريقتين تحاولان وصف العلاقة بين متغيرين (أو أكثر)؛ فإن هذين المصطلحين غالبًا ما يتم الخلط بينهما من قِبل المتخصصين وحتى العلماء. لا يضع الارتباط افتراضاً بديهياً لتبعية أحد المتغيرات لمتغيرٍ آخر من عدمه، كما لا يهتم بالعلاقة بين المتغيرات؛ ولكنه بدلًا من ذلك يُعطي تقديرًا على درجة الارتباط بين المتغيرات. وعلى الجانب الآخر؛ يحاول الانحدارُ أن يَصِف تبعية متغير الاسـتجابة لواحدٍ (أو أكثر) من المتغيرات التفسيرية؛ إذ يفترض ضمنيًّا أن هناك تأثيراً سببياً أحادي الاتجاه من المتغير التفسيري (أو المتغيرات التفسـيرية) إلى متغير الاسـتجابة، بغض النظر عمَّا إذا كان مسار التأثير مباشرًا أو غير مباشر. أيضًا وعلى الرغم من أنَّ الارتباط يهتم بالعلاقات ذات المستوى المنخفض بين متغيرين؛ فإنَّ الانحدار يختصُّ بالعلاقات بين جميع المتغيرات التفسيرية ومتغير الاستجابة.
- **الانحدار البسيط مقابل الانحدار المتعدد:** إذا تمَّ بناءُ معادلة الانحدار بين متغير استجابة واحد ومتغير تفسـيري واحد؛ عندئذٍ يُطلق عليه الانحدار البسـيط. فعلى سبيل المثال: تُعَدُّ معادلة الانحـدار التـي بُنيَت بغرض توقع/ شرح العلاقة بين طول شـخص (المتغير التفسـيري) ووزنه (متغير الاستجابة) مثالًا جيدًا على الانحدار البسيط. أمَّا الانحدار المتعدد: هو امتدادٌ للانحدار البسـيط؛ إذ تكون المتغيِّرات التفسـيرية أكثر من واحد (أو) أكثر من متغير. فعلى سبيل المثال، وفي نفس المثال السابق: إذا كنا لن نعتمد فقط على طول الشخص للتنبؤ بوزنه بل سنأخذ أيضاً في الاعتبار الخصائص الشـخصية الأخرى (مثل: مؤشر مجموعة الجسـم، الجنس، العِرْق)؛ فإننا في هذه الحالة سـنقوم بإجـراء تحليل الانحدار المتعدد. وفي كلتـا الحالتين؛ تكون العلاقة بين متغير الاسـتجابة والمتغير التفسيري (المتغيرات التفسيرية) علاقة خطية ومضافة بالطبيعة. أمَّا إذا كانت العلاقات غير خطية؛ فقد نحتاج إلى استخدام واحدٍ من العديد من طرق الانحدار غير الخطية الأخرى؛ للوصول بصورة أفضل للعلاقات بين متغيرات المدخلات والمخرجات.

كيف نطوِّر نموذج الانحدار الخطي؟

لفهم العلاقة بين متغيرين؛ فإن أبسـط شيءٍ يمكن القيام به هو رسـم مخطط تبعثر؛ إذ يمثل المحـور y (أو ص) قيـم متغير الاسـتجابة، ويمثل المحور x (أو س) قيم المتغير التفسـيري (انظر: الشـكل ٢-١٣). سـيقوم مخطط التبعثر بإظهار التغيُّرات في متغير الاسـتجابة كدالة للتغيرات في

المتغير التفســيري. يُوضِّح الشــكل ٢-١٣، وجود علاقة طردية بين الاثنين؛ فكلما زادت قيم المتغير التفسيري؛ زادت قيم متغير الاستجابة.



شكل ١٣-٢: مخطط تبعثر وخط انحدار خطي

يهدف تحليل الانحدار البســيط إلى إيجاد تمثيلٍ رياضي لهذه العلاقة. وفي الواقع؛ فهو يحاول العثور على أيٍّ دلالة لمرور خط مستقيم بشكلٍ صحيح بين النقاط المرسومة (والتي تمثل البيانات التاريخيــة/ الملاحظــة) بطريقة تقلل من المســافة بين النقاط والخط (القيــم المتوقعة على خط الانحدار النظري). وعلى الرغم من وجود عِدة طرق/ خوارزميات مقترحة لتحديد خط الانحدار؛ فإن الطريقة الأكثر شيوعًا في الاستخدام تُسمَّى طريقة المربعات الصغرى العادية (OLS). تهدف طريقة OLS إلى تخفيض مجموع المربعات المتبقية (تربيع المسافات الرأسية بين الملاحظة ونقطة الانحدار) وتؤدي إلى تعبير رياضي للقيمة التقديرية لخط الانحدار (والتي تُعرف باســم معلمات b). وبالنســبة للانحدار الخطي البسيط؛ يمكن إظهار العلاقة المذكورة أعلاه بين متغير الاستجابة (y) والمتغير التفسيري (x) كمعادلة بسيطة كما يلي:

$$y = \beta_0 + \beta_1 x$$

في هذه المعادلة؛ يُســمَّى (B0) التقاطع، ويُســمَّى (B1) الميل. وبمجرد أن تقوم OLS بتحديد قيم هذين المعاملين؛ يكون من الممكن استخدامُ المعادلة البسيطة للتنبؤ بقيم y من قيم x المعطاة. وتكشــف العلامــة وقيمة (B1) أيضًا اتجاه وقوة العلاقة بــين كلا المتغيرين. إذا كان النموذج من

نوع انحدار خطي متعدد؛ فسيستلزم الأمر تحديد مزيدٍ من المعاملات؛ بحيث يكون هناك معاملٌ واحدٌ لكلِّ متغيرٍ تفسـيري إضافي. وكما توضِّح الصيغة التالية؛ سـوف يتم ضرب المتغير التفسيري الإضافي مع معاملات (B1) الجديدة ويتم جمعها معًا؛ لإنشاء تمثيل جمعي خطي لمتغير الاستجابة.

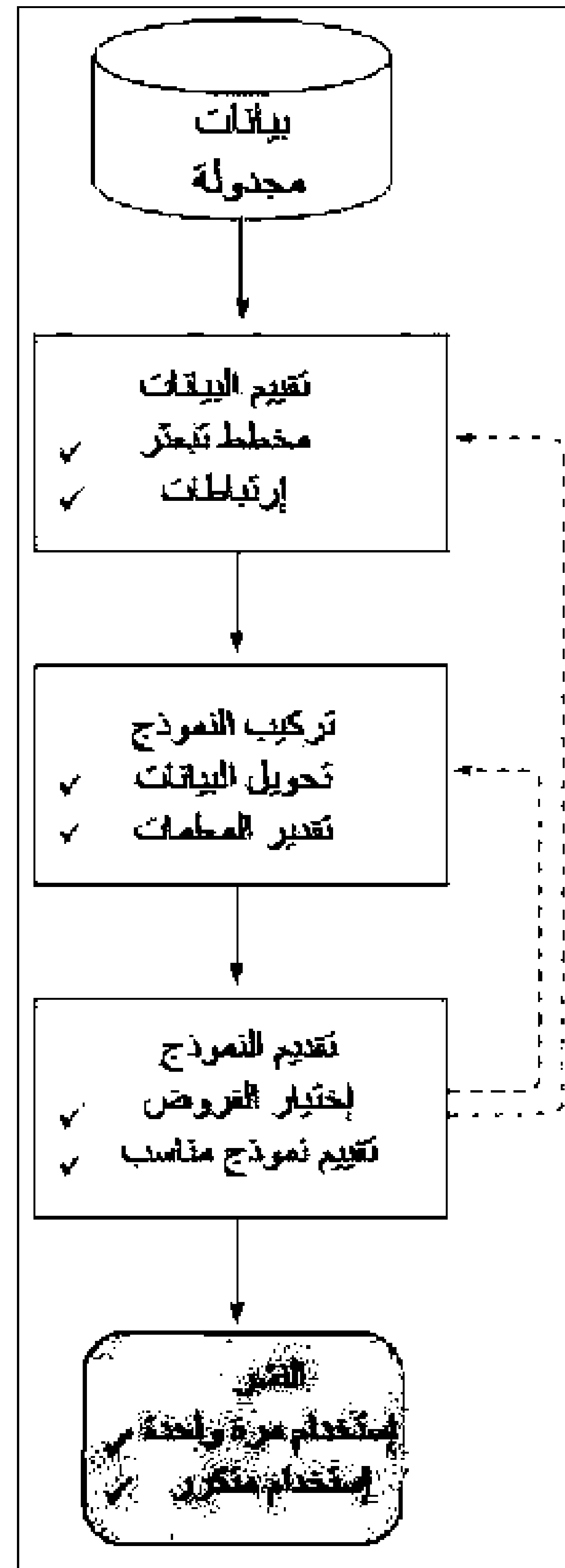
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n$$

كيف نعرف ما إذا كان النموذجُ جيدًا بما فيه الكفاية؟

لعِدة أسـبابٍ متنوعة لا تمثـل النماذج في بعض الأحيان الواقع بصـورة جيدة، وبغض النظر عن عدد المتغيرات التفسـيرية المُسـتَخدَمة في النموذج، دائماً هناك إمكانيةٌ للحصول على نماذج بجودة منخفضة، وبالتـالي يجب تقييم نموذج الانحدار الخطي وفقاً لمدى ملاءمته (الدرجة التي يمثل بها متغير الاسـتجابة). بشـكلٍ مُبسَّـط؛ ينتج عن نموذج الانحدار المناسب قيم تنبؤ تكون قريبةً من القيم الحقيقية التي تمَّت ملاحظتها في الواقع. وبالنسـبة للتقييم الرقمي؛ غالبًا ما يتمُّ اسـتخدام ثلاثة مقاييس إحصائية في تقييم ملاءمـة نموذج الانحدار. R^2 (R - تربيع)، واختبار F الشـامل، وخطأ الجذر التربيعي للوسط الحسابي (RMSE). تستندُ هذه المقاييس الثلاثة جميعًا على مجموع أخطاء المربع (كم تبعد البيانات عن الوسـط الحسابي، وكم تبعد البيانات عن القِيم المتوقعـة للنموذج). وتعمل توليفات مختلفة من هاتـين القيمتين على توفير معلومات مختلفة حول كيفية مقارنة نموذج الانحدار مع نموذج الوسط الحسابي.

من بين تلك المقاييس الثلاثة، يحتوي R^2 على المعنى الأكثر فائدةً وفهمًا بسـبب كونه مقياسًا بديهيًا. تتراوح قيمة R^2 من صفر إلى واحد (المقابل لمقدار التباين الذي يظهر كنسبة مئوية)؛ إذ يشـير "الصفر" إلى أن العلاقة وقوة التنبؤ الخاصة بالنموذج المقترح ليسـت جيدةً؛ في حين يُعبِّر «الواحد» عن أن النموذج المقترح مناسـبٌ تمامًا وينتج توقعات دقيقة (وهذا لايحدث في الواقع مطلقـاً). وعـادةً ما تكون قيم R^2 الجيـدة قريبةً من الواحد الصحيح، وهذا القرب هو أسـاس الظاهـرة التي تمَّ نمذجتها؛ بحيث إذا كانت قيمـة $R^2 = 0.3$ لنموذج الانحدار الخطي في العلوم الاجتماعيـة؛ فيمكـن اعتبار هذه القيمة جيدة بما فيه الكفاية، أمـا إذا كانت قيمة $R^2 = 0.7$ في العلوم الهندسـية؛ فهذه القيمة لا تُعَدُّ مناسـبةً بدرجة كافية. ويمكن تحقيق التحسُّن في نموذج الانحدار عن طريق إضافة متغيرات تفسيرية متسقة، أو استبعاد بعض المتغيرات خارج النموذج، أو اسـتخدام تقنيات تحويل بيانات مختلفة؛ مما يؤدي إلى زيادات نسـبية في قيمة R^2. ويوضح الشـكل ٢-١٤ عملية تدفق لتطوير نمـاذج الانحدار. وكما يتبَيَّن من عمليـة التدفق؛ فإنَّ مهمة

تطويــر النموذج لا بــد أن تتبعها مهمة تقييم النموذج؛ إذ لا يقتــصر الأمرُ على ملاءمة النموذج الذي تمَّ تقييمه فقط؛ ولكن بسبب الافتراضات التقييدية التي يتعيَّن على النماذج الخطية الالتزام بها، كما يجب وضع صلاحية النموذج تحت المجهر.



شكل ٢-١٤: عملية تدفق تطوير نماذج الانحدار

ما أهم الافتراضات في الانحدار الخطي؟

عــلى الرغم من أنــه لا يزال هناك اختيارٌ بــين العديد من تحليلات البيانات (ســواء لأغراض تفســيرية، أو لأغراض النمذجــة التنبؤية)؛ فإنَّ نماذج الانحدار الخطــي تعاني من عِدة افتراضات شديدة التقييد. وتعتمد صلاحية النموذج الخطي المبني على قدرته على الالتزام بهذه الافتراضات. ونعرض فيما يلي أكثر تلك الافتراضات وضوحًا:

١- **الخطية:** ينصُّ هذا الافتراض على أن العلاقة بين متغير الاستجابة والمتغيرات التفسيرية هي علاقةٌ خطية. مما يعني أن القيمة المتوقعة لمتغير الاستجابة؛ هي دالة خطٍّ مستقيم لكلِّ متغير تفسيري، مع تثبيت جميع المتغيرات التفسيرية الأخرى. وكذلك لا يعتمد ميلُ الخط على قيم المتغيرات الأخرى. ويعني أيضًا أنَّ تأثيرات المتغيرات التفسيرية المختلفة على القيمة المتوقعة لمتغير الاستجابة؛ هي تأثيرات مضافة بطبيعتها.

٢- **الاستقلالية (للأخطاء):** وينصُّ هذا الافتراضُ على أن أخطاء متغير الاستجابة غير مترابطة بعضها مع بعض. استقلال الأخطاء أضعف من الاستقلال الإحصائي الفعلي، وهو حالة أقوى وغالبًا لا يكون هناك لتحليل الانحدار الخطي.

٣- **الطبيعية (للأخطاء):** وينصُّ هذا الافتراضُ على أن أخطاء متغير الاستجابة تتوزَّع توزيعًا طبيعيًّا؛ أيْ إنه من المفترَض أن تكون عشوائيةً تمامًا ويجب ألا تمثل أيَّ أنماط غير عشوائية.

٤- **التباين الثابت (للأخطاء):** وهذا الافتراض، والذي يُطلَق عليه أحيانًا (التفاوت المتماثل)، ينصُّ على أن متغيرات الاستجابة يكون لها نفس التباين في خطأها، بغض النظر عن قيم المتغيرات التفسيرية. وعمليًّا؛ لا يكون هذا الافتراض صالحًا إذا اختلف متغيرُ الاستجابة على نطاق/ مقياس واسع بما فيه الكفاية.

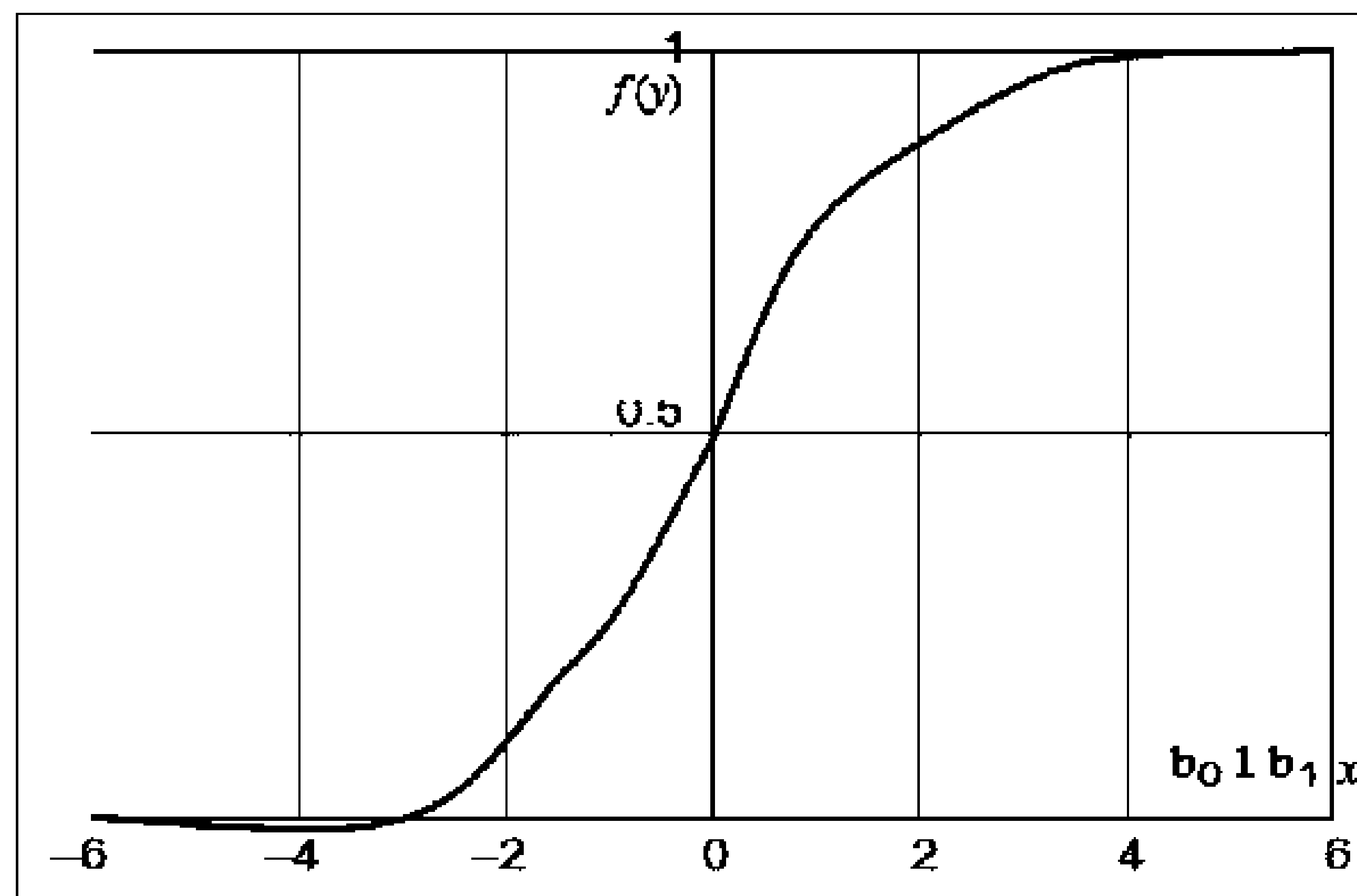
٥- **الخطية المتعدِّدة:** وينصُّ هذا الافتراضُ على عدم وجود ارتباط بين المتغيرات التفسيرية (أيْ: إنها لا تكرر نفس النتيجة؛ ولكنها تقدِّم المعلومات المطلوبة للنموذج من منظورٍ مختلف). ويمكن تطبيق الخطية المتعددة من خلال تضمين النموذج متغيرين أو أكثر من المتغيرات التفسيرية يكون بينها ارتباطٌ تام (فمثلًا: إذا تمَّ تضمينُ نفس المتغير التفسيري عن طريق الخطأ في نفس النموذج مرتين، أو مرةً واحدةً مع تحوُّل طفيف من نفس المتغير). وعادةً ما يتمُّ اكتشافُ هذا الخطأ بواسطة تقييم البيانات المعتمد على الارتباط.

هناك تقنيات إحصائية تمَّ تطويرها تعملُ على كسر هذه الافتراضات والتقنيات للتخفيف من أثرها. إن الجزء الأكثر أهميةً بالنسبة لمصمِّم النماذج هو أن يكون على دراية بوجود وسائل تقييم تلك النماذج، وأن يتمَّ وَضْع تلك الوسائل للتأكد من أن النماذج متوافقة مع الافتراضات التي يتم بناؤها وفقاً لها.

الانحدار اللوجستي:

يُعـرَّف الانحدارُ اللوجســتي على أنه خوارزميــة تصنيف، لها دلالة إحصائية، مســتندة على الاحتمالية والتي تســتخدم تعلُّماً خاضعاً للإشراف. وقد تم تطويره في أربعينيات القرن العشرين كمتممٍ لكلٍّ من طريقة تحليل الانحدار الخطي وطريقة التمايز الخطي. وقد تمَّ استخدامه على نطاقٍ واسعٍ في العديد من التخصُّصات، بما في ذلك مجالات العلوم الطبية والاجتماعية. ويتشابه الانحدار اللوجســتي مع الانحدار الخطي في كونه يهدف كذلك إلى الاعتماد على الدالة الرياضية التي تشرح العلاقة بين متغير الاســتجابة والمتغيرات التفســيرية باســتخدام عينة من الملاحظات الســابقة (بيانــات التدريب). ولكنهما يختلفان في نقطة أساســية واحــدة: وهي أن مُخرجات الانحدار اللوجستي (متغير الاستجابة) تكون فئة في حين تكون مخرجات الانحدار الخطي متغيراً رقميــاً. بمعنى أنه في الوقت الذي يُســتَخدَم فيه الانحدارُ الخطي لتقدير متغيرٍ رقمي مســتمر؛ يُســتَخدَم الانحدارُ اللوجســتي لتصنيف متغير فئــوي. وعلى الرَّغم من أن الشــكل الأصلي من الانحدار اللوجستي قد تمَّ تطويره إلى متغير مخرجات ثنائي (مثل: ١/ ٠، نعم/ لا، ناجح/ راسب، قبول/ رفض)؛ فإن النسخة المعدَّلة في الوقت الحاضر قادرةٌ على التنبؤ بمتغيرات مخرجات متعددة الفئات (أيْ: الانحدار اللوجســتي المتعدد الحدود). أمَّــا إذا كان هناك متغيرُ تنبؤ واحد، ومتغير توقُّع واحد؛ فتُســمَّى الطريقة هنا الانحدارُ اللوجســتي البســيط (وهي مشابهةٌ لطرق استدعاء نماذج الانحدار الخطي مع متغيرٍ مستقل واحد فقط كانحدار خطي بسيط).

في التحليلات التنبؤية؛ يتمُّ اســتخدامُ نماذج الانحدار اللوجســتي لتطوير نماذج احتمالية بين واحــدٍ أو أكثر من المتغيرات التفســيرية/ التوقعية (والتي قد تكون مزيجًــا من كلٍّ من الطبيعة المســتمرة والفئوية) ومتغير التصنيف/ الاســتجابة (والذي قد يكون ثنائياً/ ذا حدين أو متعدد الحــدود/ متعــدد الفئات). وعلى عكــس الانحدار الخطــي الاعتيادي؛ يتمُّ اســتخدام الانحدار اللوجستي؛ للتنبؤ بنتائج فئوية (غالبًا ما تكون ثنائية) لمتغير الاستجابة - معالجة متغير الاستجابة كنتيجــة لتجربــة Bernoulli. ولذلك؛ يقــوم الانحدارُ اللوجســتي بأخــذ اللوغاريتم الطبيعي لاحتمالات متغير الاســتجابة لإنشاء معيار مستمر كنسخة محوَّلة لمتغير الاستجابة. وبالتالي يُشار إلى تحويل logit على أنه دالة الربط في الانحدار اللوجستي - وعلى الرغم من أن متغير الاستجابة في الانحدار اللوجســتي؛ هو متغيرٌ فئوي أو ثنائي؛ فإن logit هو المعيارُ المســتمر الذي يتم على أساسه الانحدار الخطي. ويوضِّح الشكل ٢-١٥ دالةَ الانحدار اللوجستي؛ إذ يتمُّ تمثيلُ الاحتمالات في المحور x (أو س) (دالة خطية للمتغيرات المستقلة)، في حين تظهر النتيجة الاحتمالية في المحور y (أو ص) (أيْ: تتراوح قيم متغير الاستجابة بين ٠ و١).



شكل ٢-١٥: الدالة اللوجستية

تُعَدُّ الدالة اللوجستية، f(y) في الشكل ١٥-٢، هي صميم الانحدار اللوجستي، والذي لا يمكن أن يأخذ سوى القيم بين الصفر والواحد الصحيح. وتُعَدُّ المعادلة التالية تمثيلًا رياضيًّا مبسطًا لهذه الدالة:

$$f(y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

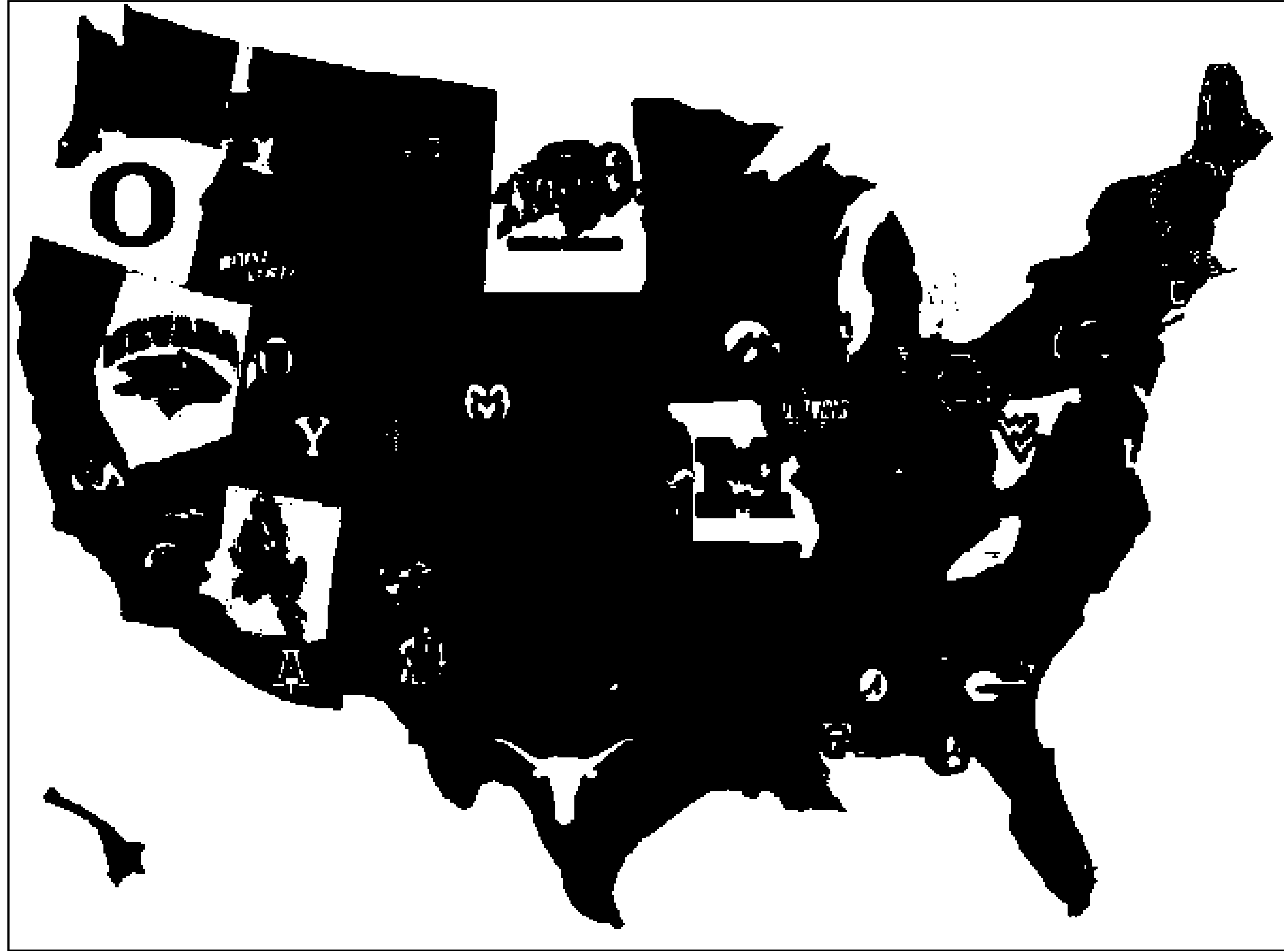
يتمُّ عادةً تقديرُ معاملات الانحدار اللوجستي (the bs) باستخدام الطريقة القصوى لتقدير الاحتمالية. وعلى عكس الانحدار الخطي مع البواقي الموزَّعة توزيعًا طبيعيًّا؛ فإنه ليس من الممكن العثورُ على مصطلح مغلق لقيم المُعامل التي تعظم دالة الاحتمال؛ ولذا ينبغي استخدامُ عملية تكرارية بدلًا من ذلك. تبدأ هذه العملية بحلٍّ مبدئي مؤقت، ثم تقوم بمراجعة المعلمات قليلًا لمعرفة ما إذا كان يمكِن تحسينُ الحل وإعادة هذه المراجعة التكرارية حتى لا يمكن تحقيق أيِّ تحسُّن أو الوصول للحد الأدنى للغاية، وعند هذه النقطة يُقال: إنَّ العملية قد اكتملت/ تقاربت.

كانت التحليلاتُ الرياضية تحظى بشعبية هائلة. وأصبح استخدامُ تقنيات التحليلات المبنية على البيانات اتجاهًا سائدًا ليس فقط للفِرَق الاحترافية؛ بل أيضًا للرياضات الجامعية والهواة. والتحليلات الرياضية تعني استخدام البيانات والتقنيات الإحصائية/ التحليلية لإدارة الفِرَق/ المنظمات الرياضية على نحوٍ أفضل. وتُعَدُّ الحالة العملية ٢-٤ مثالًا على كيفية استخدام مصادر البيانات العامة الموجودة والمتاحة بسهولة للتنبؤ بنتائج مباراة في كرة القدم الجامعية باستخدام كلٍّ من نماذج التصنيف ونماذج التنبؤ بنوع الانحدار.

حالة عملية ٢-٤

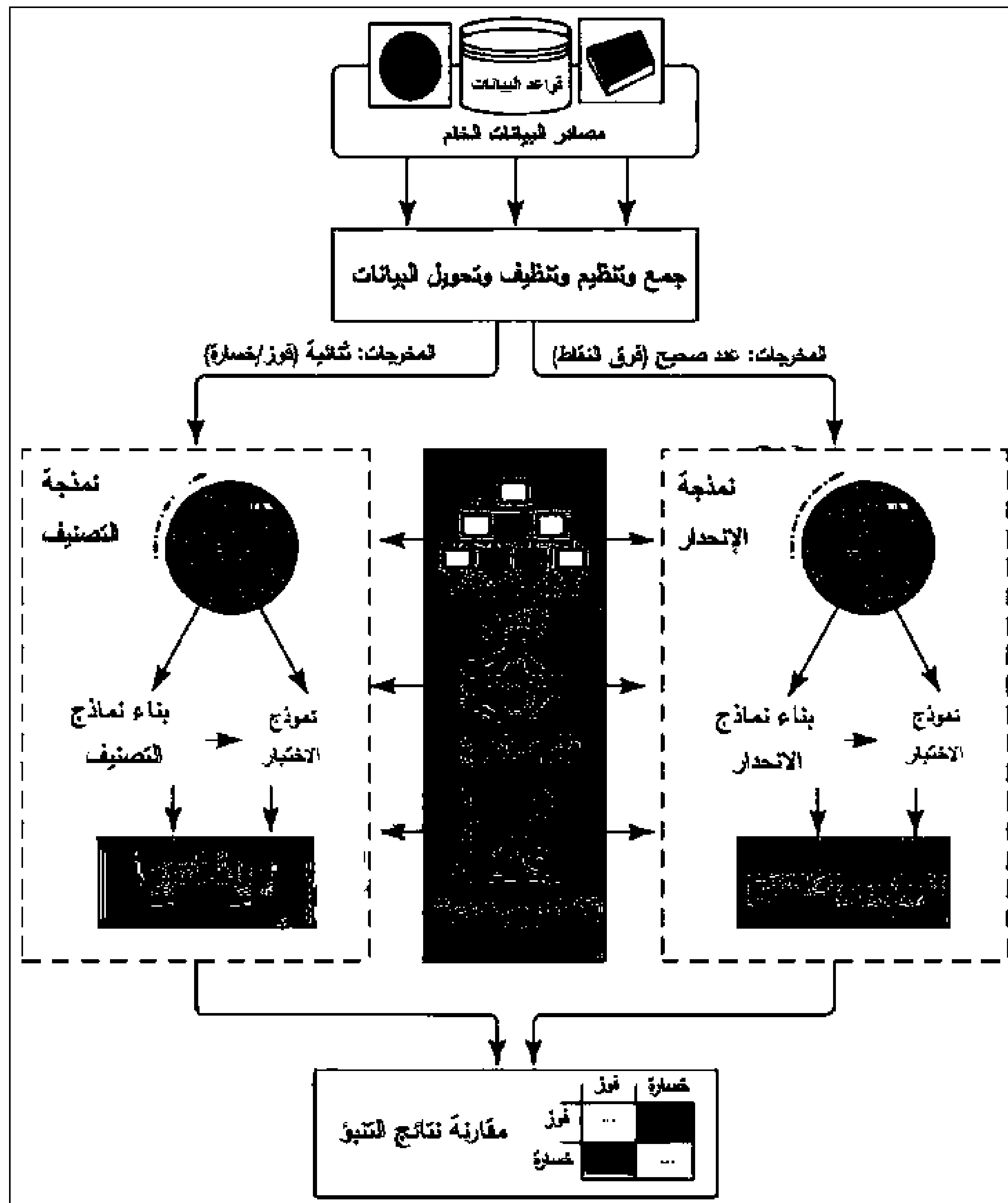
توقع نتائج لعبة وعاء NCAA

يُعَدُّ التنبؤ بنتيجة لعبة كرة القدم الجامعية (أو أيّ لعبة رياضية أخرى) مشكلةً مثيرةً للاهتمام ومليئةً بالتحَدِّيات. ولذلك؛ فقد بذل هواة التحَدِّي من الباحثين سواءً الأكاديميين أو الصناعيين جهدًا عظيمًا في التنبؤ بنتائج الأحداث الرياضية. تُوجَد كمياتٌ كبيرةٌ من البيانات التاريخية في منافذ إعلامية مختلفة (غالبًا ما تكون معلنةً) فيما يتعلَّق بهيكل ونتائج الأحداث الرياضية في شكلِ مجموعة متنوعة من العوامل التي يتمُّ تمثيلها عدديًا أو رمزيًا والتي يفترض أنها تُسهِم في تلك النتائج. يُعَدُّ وعاء ألعاب نهاية الموسم في غاية الأهمية بالنسبة للكليات من الناحية المالية (لتحقيق عائدٍ إضافي يصل لملايين الدولارات) بالإضافة إلى السُّمعة؛ وذلك بغرض ضمِّ الطلاب المتفوقين والرياضيين أصحاب التقديرات الكبيرة على برامجهم الرياضية في أثناء مرحلة الدراسة الثانوية (Freeman & Brewer, 2016). تقوم الفرق التي يتمُّ اختيارها للتنافس في وعاء لعبة معينة باقتسام الجائزة المالية، بمقدارٍ يعتمد على ذلك الوعاء لكلٍّ منهم (بعض الأوعية تكون أكثر فخامةً وذات دفعات أعلى للفريقين)؛ ولذلك يُعَدُّ ضمان دعوة إلى وعاء اللعبة؛ هو الهدف الرئيسي لأيِّ برنامج كرة قدم لقسم I-A بالكلية. ويتمُّ مَنْح متخذي القرارات لألعاب الوعاء صلاحيةَ اختيار ودعوة الجمهور المستحق (فريق لديه ستة انتصارات ضد خصومه في القسم I-A في ذلك الموسم) للفِرَق الناجحة (حسب التقديرات والتصنيفات العالمية) التي ستلعب المباراة بشكلٍ مثيرٍ وتنافسي، مع الحفاظ على ما تبقى من المشجِّعين المنضبطين عبر منافذ الإعلام المتنوعة. وفي دراسة حديثة حول التنقيب في البيانات؛ استخدم كلٌّ من Delen وCogdell وKasap ثماني سنوات من بيانات لعبة bowl إلى جانب ثلاث تقنيات شائعة للتنقيب في البيانات (أشجار القرار، والشبكات العصبية، وآلة المتجهات الداعمة)؛ للتنبؤ بكلٍّ من نتائج تصنيف إحدى الألعاب (الفوز مقابل الخسارة)؛ فضلًا عن النتيجة من نوع الانحدار (فِرق النقاط المسقطة بين درجات الخصمين). وفيما يلي وصفٌ مختصرٌ لدراستهم تلك.



المنهجية:

في هــذا البحث؛ اتبع دلين وزملاؤه منهجيةً واســعة الانتشــار في التنقيب في البيانات تُســمَّى CRISP-DM (عملية قياســية عبر الصناعــة لتعدين البيانات)، وهي عمليةٌ من ســت خطوات. وقد وفَّرَت هذه المنهجية المشهورة، والتي سيتمُّ تناولها بالتفصيل في الفصل الرابع، طريقةً منهجيةً ومنظمة لإجراء دراسة التنقيب في البيانــات الأساســية، ومن ثــم زادت من احتمالية الحصول عــلى نتائج دقيقة وموثوقة. ولتقييم قوة التنبؤ لأنواع النماذج المختلفة بموضوعية؛ قاموا باستخدام منهجية للتحقق من الفاعلية، والتي تُسمَّى س-أجزاء (K-folds). ويمكن الاطلاع عــلى تفاصيل منهجيــة k-fold في الفصل الرابــع. ويوضِّح الشــكل ٢-١٦ بيانيًّا المنهجية المُستَخدَمة من قِبل الباحثين.



شكل ٢-١٦: الرسم التوضيحي للمنهجية المستخدمة في الدراسة

اقتناء ومعالجة البيانات:

تمَّ جمعُ بيانات العينة لهذه الدراسة من مجموعة متنوعة من قواعد البيانات الرياضية المتاحة على الويب، بما في ذلك jhowel.net وESPN.com وCovers.com وncaa.org وrauzulusstreet.com. تضمَّنت مجموعة البيانات ٢٤٤ مباراةً، تمثل مجموعةً كاملةً من ثمانية مواسم من مباريات كرة القدم الجامعية التي تمَّ لعبها بين عامي ٢٠٠٢ و٢٠٠٩م. وقد قمنا أيضًا بتضمين مجموعة بيانات خارج العينة (ألعاب ٢٠١٠-٢٠١١م)؛ للحصول على أغراض فاعلية إضافية. وقد

تمَّ استخدامُ بديهية واحدة من البديهيات الرائجة في التنقيب في البيانات، وقاموا بتضمين الكثـير من المعلومات ذات الصِّلة في النموذج قدر الإمكان. لذلك؛ فبعد عمليـة متعمقة لتحديد وتجميـع المتغير، انتهى بهم الأمـرُ إلى مجموعة بيانات شـملت ٣٦ متغيرًا، كانت أول ٦ متغيرات منها تتعلَّق بتحديد الهوية (أيْ: اسـم وتاريخ المباراة، وأسـماء الفِرق، صاحب الأرض والضيـف، ومؤتمراتهم الرياضية - انظـر: المتغيرات من ١ إلى ٦ في الجدول ٢-٥)، ثـم يتبعهم ٢٨ متغيرًا للمدخلات (تضمَّنَت متغيرات لتحديد الإحصاءات الموسـمية للفريق بشأن الهجوم والدفاع، ونتائج المباريات، وخصائص تكوين الفريق، وخصائص المؤتمرات الرياضية، وكيف نجحوا في تخَطِّي الصعاب - انظر: المتغيرات من ٧ إلى ٣٤ في الجدول ٢-٥)، وأخيرًا؛ كان آخـرُ اثنين من المتغيرات للمخرجات (وهما ScoreDiff وWinLoss - حيث يعنـي ScoreDiff فرق النقاط بين الفريق صاحب الأرض والفريق الضيف ممثلةً بعددٍ صحيح، في حين يشـير WinLoss إلى فوز أو خسـارة الفريق صاحب الأرض للعبة الوعاء والممثلة بعلامة اسمية).

في صياغـة مجموعة البيانات؛ قـام كل صـفٍّ (a.k.a. tuple، الحالة، العينة، المثال... إلخ) بتمثيل لعبة وعاء، واختص كلُّ عمودٍ بمتغير (أيْ: نوع المتغير معرف/ مدخـلات أو مخرجات). ولتمثيل الخصائص النسـبية المرتبطة باللعبة للفريقين المتنافسـين، في متغيرات المدخلات؛ قمنا بحسـاب واستخدام الفروق بين مقاييس الفريـق صاحب الأرض والفريق الضيف. ويتمُّ حسـابُ كلِّ قيم المتغير هذه من منظور الفريق صاحب الأرض. على سبيل المثال: يمثل المتغير PPG (متوسط عدد النقـاط التي سـجلها الفريق في كل مباراة) الفرق بين الـ PPG للفريق صاحب الأرض والـ PPG للفريـق الضيف. وتمثِّـل متغيرات المخرجات فوز أو خسـارة الفريـق صاحب الأرض للمباراة. بمعنى، لو يأخذ متغير ScoreDiff رقمًا صحيحًا موجبًا؛ فمـن المتوقع أن يفوز الفريق صاحـب الأرض باللعبة بهذا الهامش، وإلا (أيْ: لـو يأخذ المتغير ScoreDiff رقمًا صحيحًا سـالبًا) فعندئـذٍ نتوقع أن يخسر الفريق صاحب الأرض اللعبة بهذا الهامش. في حالة WinLoss، تكون قيمة متغير المخرجـات عبارة عن تسـمية ثنائية، "Win" أو "Loss" وهي تشـير إلى نتيجة المباراة بالنسبة للفريق صاحب الأرض.

النتائج والتقييم:

في هذه الدراسة؛ يتمُّ استخدامُ ثلاث تقنيات تنبؤ رائجة لبناء النماذج (ومقارنتها بعضها مع بعض): الشبكات العصبية الاصطناعية، وأشجار القرار، وآلة المتجهات الداعمة. ويتمُّ اختيار تقنيات التنبؤ هذه وفقاً لقدرتها على نمذجة كلٍّ من التصنيفات إضافةً إلى مشكلات التنبؤ من النوع الانحداري وانتشارها في أدبيات التنقيب في البيانات المنشورة حديثًا. ويمكن العثور على مزيدٍ من التفاصيل حول طرق التنقيب في البيانات الشائعة هذه في الفصل الرابع.

لمقارنة الدقة التنبؤية لجميع النماذج بعضها ببعض؛ استخدم الباحثون منهجية التحقق المتقاطع الطبقي k-fold. في النسخة الطبقية من k-fold، يتمُّ إنشاء الطبقات بطريقة تحتوي على نفس النسبة تقريبًا من علامات التنبؤ (أيْ: الطبقات) كما في مجموعة البيانات الأصلية. وفي هذه الدراسة؛ يتمُّ وَضْع قيمة k = 10 (أيْ: يتم تقسيم المجموعة الكاملة المكوَّنة من ٢٤٤ عينة إلى ١٠ مجموعات فرعية، بكلٍّ منها ٢٥ عينة تقريبًا)، والذي يُعَدُّ ممارسةً شائعةً في تطبيقات التنقيب عن البيانات التنبؤية. وقد سبق توضيح k-fold برسمٍ بياني قبل ذلك في هذا الفصل. ولمقارنة نماذج التنبؤ التي تمَّ تطويرها باستخدام تقنيات التنقيب الثلاث المذكورة آنفًا؛ فقد استقرَّ الباحثون على استخدام ثلاثة معايير أداء شائعة الاستخدام، هي: الدقة، والحساسية، والخصوصية. كما تمَّ شرح المعادلات البسيطة لهذه المقاييس في وقتٍ سابق في هذا الفصل.

جدول ٢-٥: وصف للمتغيرات المستخدمة في الدراسة

م	الفئة	إسم المتغير	الوصف
1	ID	YEAR	سنة لعبة الوعاء
2	ID	BOWLGAME	اسم لعبة الوعاء
3	ID	HOMETEAM	الفريق صاحب الأرض (كما هو مُدرَج من قِبل منظمي الوعاء)
4	ID	AWAYTEAM	الفريق الضيف (كما هو مُدرَج من قِبل منظمي الوعاء)
5	ID	HOMECONFERENCE	مؤتمر الفريق صاحب الأرض

م	الفئة	إسم المتغير	الوصف
6	ID	AWAYCONFERENCE	مؤتمر الفريق الضيف
7	I1	DEFPTPGM	نقاط دفاعية لكلِّ مباراة
8	I1	DEFRYDPGM	ساحات الاندفاع الدفاعية لكلِّ مباراة
9	I1	DEFYDPGM	ساحات دفاعية لكلِّ مباراة
10	I1	PPG	متوسط عدد النقاط التي سجلها فريقٌ معين في كلِّ مباراة
11	I1	PYDPGM	متوسط عدد التمريرات الميدانية الإجمالي لكلِّ مباراة
12	I1	RYDPGM	متوسط إجمالي عدد مرات الفوز في كلِّ فريق لكلِّ مباراة
13	I1	YRDPGM	متوسط عدد الهجمات الميدانية لكل مباراة
14	I2	٪HMWIN	نسبة الفوز الرئيسية
15	I2	LAST7	كم عدد المباريات التي فاز بها الفريق من آخر ٧ مباريات
16	I2	MARGOVIC	متوسط هامش الفوز
17	I2	NCTW	نسبة فوز فريق Nonconference
18	I2	PREVAPP	هل ظهر الفريق في مباراة في العام الماضي
19	I2	٪RDWIN	نسبة الفوز على الطريق
20	I2	SEASTW	نسبة الفوز للسنة
21	I2	TOP25	الفوز نسبة ضد أعلى ٢٥ فريق AP للسنة
22	I3	TSOS	قوة الجدول الزمني للسنة
23	I3	٪FR	نسبة المباريات التي يلعبها لاعبو الصف الأول في السنة
24	I3	٪SO	نسبة المباريات التي يلعبها لاعبو الصف الثاني في السنة
25	I3	٪JR	نسبة المباريات التي يلعبها اللاعبون الأحدث سنًّا لهذا العام
26	I3	٪SR	النسبة المئوية للألعاب التي لعبها لاعبون من الدرجة الأولى خلال العام
27	I4	٪SEASOvUn	نسبة المرات التي مَرَّ فيها الفريق على O/U* في الموسم الحالي
28	I4	٪ATSCOV	مقابل نسبة تغطية انتشار الفريق في المباريات السابقة
39	I4	٪UNDER	النسبة المئوية لمرات انهيار الفريق في المباريات السابقة
30	I4	٪OVER	نسبة المرات التي اجتازها الفريق في مباريات سابقة
31	I4	٪SEASATS	نسبة التغطية مقابل الانتشار للموسم الحالي

م	الفئة	إسم المتغير	الوصف
32	I5	CONCH	هل فاز الفريق ببطولة المؤتمر الخاصة بهم؟
33	I5	CONFSOS	قوة المؤتمر من الجدول الزمني
34	I5	٪CONFWIN	نسبة الفوز في المؤتمر
35	O1	ScoreDiffo	فرق النقاط (نقاط صاحب الأرض - نقاط الضيف)
36	O2	WinLosso	ما إذا كان الفريق صاحب الأرض يفوز أو يخسر اللعبة

* O/U - ما إذا كان الفريق سيخترق أو يقل عن فرق النقاط المتوقعة.

o متغيرات الإخراج: ScoreDiff لنماذج الانحدار وWinLoss لنماذج التصنيف الثنائية.

I1: المخالفة/ الدفاع؛ I2: نتيجة اللعبة؛ I3: تكوين الفريق؛ I4: ضد الصعاب؛ I5: إحصائيات المؤتمر.

ID: متغيرات تحديد الهوية؛ O1: متغير المخرجات لنماذج الانحدار؛ O2: متغير المخرجات لنماذج التصنيف.

ويتــمُّ تقديمُ نتائج التنبؤ لتقنيــات النمذجة الثلاث في الجدول ٢-٦ والجدول ٢-٧؛ إذ يعــرض الجــدول ٢-٦ نتائج الفاعلية المتقاطعة مــن ١٠ أضعاف لمنهجية التصنيــف؛ إذ تتمُّ صياغــة تقنيات التنقيب في البيانــات الثلاثة؛ بحيث يكون لها متغــير مُخرجــات ثنائي التســمية (أيْ: WinLoss). في حين يُقــدِّم الجدول ٢-٧ نتائج الفاعلية المتقاطعة من ١٠ أضعاف لمنهجية التصنيف المستندة إلى الانحدار؛ إذ تُصــاغ تقنيــات التنقيب في البيانات الثلاثة؛ بحيث يكــون لها متغير مخرجات رقمــي (أيْ: ScoreDiff). في التنبؤ بالتصنيف القائــم على الانحدار؛ يتمُّ تحويلُ المخرجات العددية للنماذج إلى تصنيفات عن طريق وَضْع علامة "Win" على أرقام WinLoss الموجبة وعلامة "Loss" على أرقام WinLoss السالبة، ثم جدولة ذلك كله في مصفوفات الارتباك. باستخدام مصفوفات الارتباك؛ يتمُّ حساب دقة التنبؤ الشــاملة، والحساســية، والنوعية لكلٍّ نوعٍ من أنواع النــماذج وعرضها في هذين الجدولين. وكما تشــير النتائج؛ فإن أساليب التنبؤ من نوع التصنيف تؤدي بصورة أفضل من منهجية التنبؤ من نوع التصنيف القائم على الانحدار. ومن بين تقنيات التنقيب في البيانات الثلاثة؛ أنتجت أشــجارُ التصنيف والانحدار دقةً تنبؤيةً أفضل في كلٍّ من منهجيتي التنبؤ المُشــار إليهما. وبشــكلٍ عام؛ أنتجت نماذج التصنيف ونماذج شجرة الانحدار لمنهجية الفعالية المتقاطعة ١٠ أضعاف نسبة دقة تصل إلى ٨٦,٤٨٪، تليها آلة المتجهات الداعمة بنســبة دقة تصل إلى ٧٩,٥١٪، ثم الشبكات العصبية بنســبة دقة تصل إلى ٧٥٪. وباســتخدام اختبار t، وجد الباحثون أنَّ قيم

الدقة هذه كانت مختلفةً بشكلٍ معنوي عند مستوى ألفا ٠,٠٥، أي إنَّ شجرة القرارات تُعَد مؤشِّرًا أفضل بشكلٍ معنوي لهذا النطاق من الشبكة العصبية وآلة نقل الدعم، وأيضًا آلة نقل الدعم تُعَد مؤشرًا أفضل بكثير من الشبكات العصبية.

جدول ٢-٦: نتائج التنبؤ لمنهجية التصنيف المباشر

طريقة التنبؤ (تصنيف*)		مصفوفة الارتباك		الدقة ** (٪)	الحساسية (٪)	الخصوصية (٪)
		فوز	خسارة			
ANN (MLP)	فوز	٩٢	٤٢	٧٥,٠٠	٦٨,٦٦	٨٢,٧٣
	خسارة	١٩	٩١			
SVM (RBF)	فوز	١٠٥	٢٩	٧٩,٥١	٧٨,٣٦	٨٠,٩١
	خسارة	٢١	٨٩			
DT (C&RT)	فوز	١١٣	٢١	٨٦,٤٨	٨٤,٣٣	٨٩,٠٩
	خسارة	١٢	٩٨			

* متغير المخرجات: هو متغير فئوي ثنائي (Win أو Loss)؛ كانت الفروق معنوية.

** P < 0.01.

جدول ٢-٧: نتائج التنبؤ لمنهجية التصنيف القائم على الانحدار

طريقة التنبؤ (قائم على الانحدار*)		مصفوفة الارتباك		الدقة **	الحساسية	الخصوصية
		فوز	خسارة			
ANN (MLP)	فوز	٩٤	٤٠	٧٢,٥٤	٧٠,١٥	٧٥,٤٥
	خسارة	٢٧	٨٣			
SVM (RBF)	فوز	١٠٠	٣٤	٧٤,٥٩	٧٤,٦٣	٧٤,٥٥
	خسارة	٢٨	٨٢			
DT (C&RT)	فوز	١٠٦	٢٨	٧٧,٨٧	٧٦,٣٦	٧٩,١٠
	خسارة	٢٦	٨٤			

* متغير المخرجات هو متغير رقمي/ عدد صحيح (فرق النقط)؛ كانت الفروق معنوية.

** P < 0.01.

وقـد أظهرت نتائجُ الدراسـة أنَّ نماذج التصنيف النوعية تتنبـأ بنتائج اللعبة بشـكلٍ أفضل من نماذج التصنيف القائمة على الانحدار. وعلى الرغم من أنَّ هذه النتائـج خاصة بالنطاق العملي والبيانات المسـتَخدَمة في هذه الدراسـة، وبالتالي لا ينبغـي تعميمها خارج نطاق الدراسـة؛ فهـي مثيرةٌ؛ وذلك لأن أشـجار القرار ليسـت الأفضل في التنبؤ فحسب ولكن أيضًا هي الأفضل في الفهم والنشر، مقارنةً مع التقنيتين الأخرتين لتعلُّم الآلة المسـتخدمتين في هذه الدراسـة. ويمكن العثور على مزيدٍ من التفاصيل حول هذه الدراسة في (Delen وآخرون ٢٠١٢).

أسئلة للمناقشة:

١- ما هـي التحديات المتوقعـة في التنبؤ بنتائج الأحداث الرياضية (على سـبيل المثال: ألعاب مباراة الكلية)؟

٢- كيـف قـام الباحثون بصياغة/ تصميم مشـكلة التنبـؤ (أيْ: ما هي المدخلات والمخرجات؟ وماذا كان تمثيل عينة واحدة - صف بيانات؟)

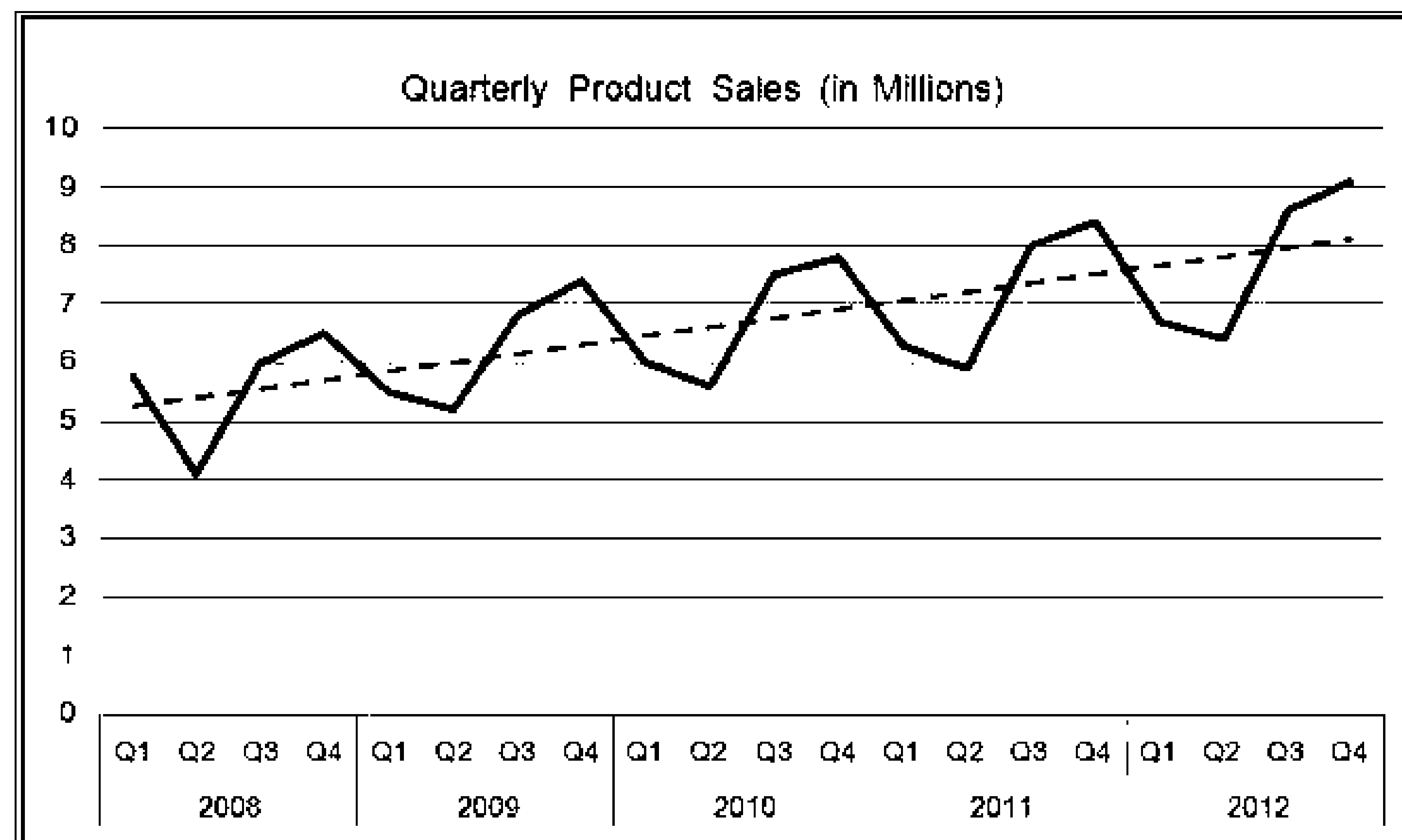
٣- ما مدى نجاح نتائج التنبؤ؟ وما الذي يمكنهم فعله لتحسين الدقة؟

Sources: Delen, D., Cogdell, D., & Kasap, N. (2012). A compara- tive analysis of data mining methods in predicting NCAA bowl outcomes. International Journal of Forecasting, 28, 543-552; Freeman, K. M., & Brewer, R. M. (2016). The politics of American college football. Journal of Applied Business and Economics, 18(2), 97-101.

التنبؤ بالسلاسل الزمنية:

في بعض الأحيان قد لا يكون للمتغير الذي نحن بصدد دراسته (أيْ: متغير الاستجابة) متغيرات تفسـيرية واضحـة المعالم، أو قد يكـون هناك الكثيرُ منها في علاقة مُعقـدة للغاية. في مثل هذه الحالات؛ إذا كانت البيانات متاحةً في الشكل المطلوب، فيمكن تطوير نموذج التنبؤ، والذي يُسمَّى السلسـلة الزمنية. تُعرَّف السلسـلة الزمنية على أنها سلسـلةٌ من نقاط البيانـات للمتغيِّر محل الاهتمام، ويتم قياسها وتمثيلها في نقاطٍ متتالية في الوقت متباعدة في فواصل زمنية مُوحَّدة. ومن أمثلة السلاسـل الزمنية أحجام الأمطار الشـهرية في منطقة جغرافية مـا، وقيمة الإغلاق اليومي لمؤشرات سـوق الأوراق المالية، ومجموع المبيعات اليومية لمحل بقالة. وفي كثيرٍ من الأحيان؛ يتمُّ تصوير السلسـلة الزمنية باسـتخدام منحنى خطي. ويُوضِّح الشـكل ٢-١٧ مثالًا لسلسلة زمنية لحجم المبيعات للسنوات من ٢٠٠٨ حتى ٢٠١٢ على أساس ربع سنوي.

يُعرف التنبؤ بالسلســلة الزمنية على أنه استخدامُ النمذجة الرياضية للتنبؤ بالقيم المستقبلية للمتغير محل الاهتمام بناءً على القِيم التي تمَّ رصدها من قبل. تبدو مخططات/ جداول السلاسل الزمنية مشابهةً جداً للانحدار الخطي البسيط في ذلك كما كان الحال في الانحدار الخطي البسيط، في السلســلة الزمنية يُوجَد متغيران: متغير الاستجابة، ومتغير الوقت المعروض في مخطط مبعثر. وبخلاف هذا التشــابه؛ لا يكاد يكون هناك أيُّ قواسم مشتركة أخرى بين الاثنين. وعلى الرغم من أنَّ تحليــل الانحدار يتمُّ توظيفه في اختبار النظريات؛ لمعرفة ما إذا كانت القيمُ الحالية لواحدٍ أو أكثر من المتغيرات التفسيرية تفسر (وبالتالي تتنبأ بـ) متغير الاستجابة؛ فإن نماذج السلاسل الزمنية تركِّز على الاســتقراء على ســلوكها المتغير مع الوقــت لتقدير القيم المســتقبلية. ويفترض التنبؤ بالسلاسل الزمنية أن جميع المتغيرات التفسيرية يتمُّ تجميعها واستهلاكها في دراسة سلوك اختلاف الوقت لمتغير الاســتجابة. ولذلك؛ فإن اكتشــاف ســلوك تغيُّر الوقت؛ هو طريقــة للتنبؤ بالقيم المســتقبلية لمتغير الاســتجابة. وللقيام بذلك؛ يتمُّ اختصارُ النمط وتحليله إلى مكوناته الرئيسية: الاختلافات العشــوائية، والاتجاهات الزمنية، والدورات الموسمية. ويوضِّح مثال السلاسل الزمنية المعروض في الشكل ٢-١٧ كلَّ هذه الأنماط المميزة.



شكل ٢-١٧: سلسلة زمنية للبيانات عن أحجام المبيعات ربع السنوية

وتتراوح التقنيات المُســتَخدَمة في تطوير تنبؤات التسلســل الزمني من بسيطة جدًا (مثل التنبؤ البســيط الــذي يقترح أن يكون الفعلي بالأمــس هو المتوقع اليوم) إلى معقــد جدًا مثل ARIMA (وهي طريقةٌ تجمع بين أنماط المتوســطات المنحدرة والمتحركة في البيانات). وربما تكون التقنيات الأكثر شيوعًا؛ هي طرقُ إيجاد المعدل، والتي تتضمَّن المتوسط البسيط والمتوسط المتحرك، والمتوسط

المتحـــرك المرجـــح، والتضخيم الأسي. كما أنَّ العديدَ من هذه التقنيات لها إصداراتٌ متقدمةٌ بحيث يمكنها أيضًا أن تأخذ في حُسبانها كلًا من الموسمية والاتجاه للتنبؤ بشكلٍ أفضل وأكثر دقة. وعادةً ما يتمُّ تقييم دقة الطريقة من خلال حساب خطأها (أيْ: الانحراف المحسوب بين الفعليات والتنبؤات الخاصة بالملاحظات السابقة) عن طريق الوسط الحسابي للخطأ المطلق (MAE)، أو الوسط الحسابي للخطأ التربيعي (MSE) أو الوسط الحسابي للخطأ المطلق في النسبة المئوية (MAPE). وعلى الرغم من أن جميعهم يســتخدمون نفس مقياس الخطأ الأســاسي؛ فإنَّ طرق التقييــم الثلاث هذه تؤكد الجوانب المختلفة للخطأ؛ إذ تقوم بتركيب بعض الأخطاء الكبيرة أكثر من الأخطاء الأخرى.

أسئلة مراجعة على القسم ٢-٦:

١- ما هو الانحدار؟ وما هو الغرض الإحصائي الذي يخدمه؟

٢- ما هي أوجه الشبه والاختلاف بين الانحدار والارتباط؟

٣- ما هو OLS؟ وكيف يحدد OLS خط الانحدار الخطي؟

٤- اذكر مع الشرح الخطوات الرئيسة لمتابعة تطوير نموذج الانحدار الخطي.

٥- ما هي أكثر الافتراضات شيوعًا للانحدار الخطي؟

٦- ما هو الانحدار اللوجستي؟ وكيف يختلف عن الانحدار الخطي؟

٧- ما هي السلاسل الزمنية؟ وما هي تقنيات التنبؤ الرئيسة لبيانات السلاسل الزمنية؟

٢-٧ إعداد تقارير الأعمال:

يحتاج صانعو القرار إلى معلوماتٍ لاتخاذ قرارات دقيقة في الوقت المناسب؛ إذ تُعَدُّ المعلومات في الأســاس ســياق البيانات. إضافةً إلى الوســائل الإحصائية التي تمَّ شرحها في القســم السابق، ويمكن أيضًا الحصول على المعلومات (التحليلات الوصفية) باســتخدام أنظمة معالجة التحليلات عبر الإنترنت [OLTP] (انظر: التصنيف البســيط للتحليل الوصفي في الشكل ٢-٧). عادةً ما يتمُّ تقديــم المعلومات إلى صُناع القرار في شــكل تقريرٍ مكتوب (رقميًــا أو ورقيًا)، مع وجود إمكانية لتقديمه شفويًا أيضًا. وببساطة؛ فإن أيَّ تقرير هو عبارة عن أيِّ وسيلة اتصال تمَّ إعدادها؛ بغرض نقل المعلومات في شكلٍ مقبول إلى كلِّ مَنْ يحتاج إليها، في أيِّ وقت وفي أيِّ مكان قد يحتاج إليها فيه. وعادةً ما يكون التقريرُ عبارة عن مســتندٍ يحتوي على معلومات (عادةً ما تكون مبنيةً على البيانات) منظمةً في شكل سرد و/ أو بياني و/ أو مجدول؛ يتمُّ إعدادها بشكلٍ دوري (متناوب) أو

على أساس الحاجة (حسب الغرض)، بالإشارة إلى فترات زمنية مُحدَّدة، أو أحداث أو حوادث أو موضوعات. وبإمكان تقارير الأعمال القيام بالعديد من الوظائف المختلفة (ولكنها غالبًا ما تكون ذات صِلة بعضها ببعض). وفي ما يلي أمثلةٌ على أكثر هذه التقارير انتشارًا:

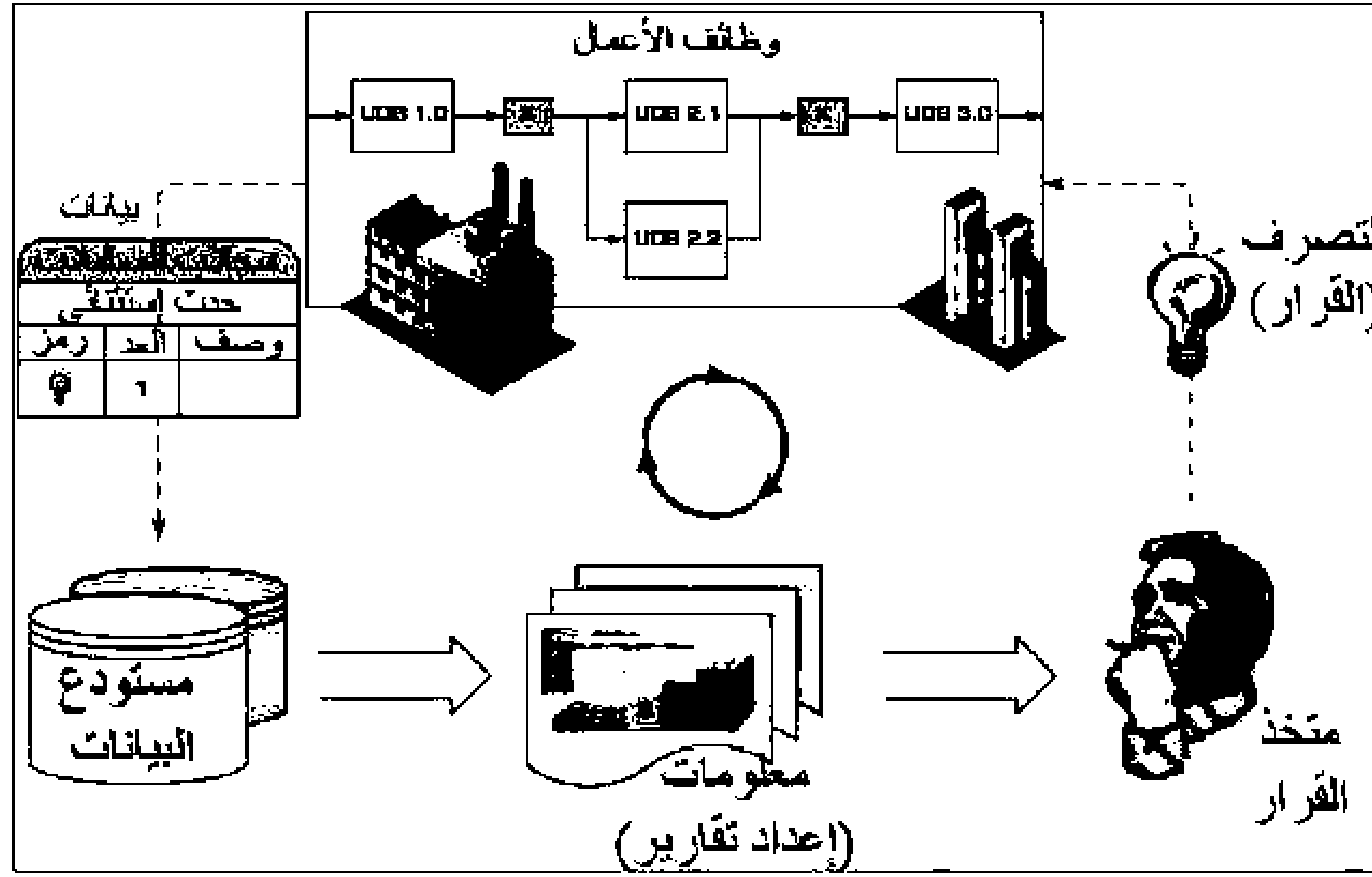
- تقارير التأكُّد من قيام جميع الإدارات بوظائفها بشكلٍ صحيح.
- تقارير لتوفير المعلومات.
- تقارير لتوفير نتائج التحليل.
- تقارير لحَثِّ الآخرين على الأداء.
- تقارير إنشاء ذاكرة تنظيمية (كجزء من نظام إدارة المعرفة).

يُعَدُّ إعداد تقارير الأعمال (والذي يُسمَّى أيضًا OLAP أو ذكاء الأعمال) جزءًا أساسيًّا من أدوات القيادة العليا بشأن اتخاذ القرارات الإدارية المثالية المُحسَّنة والقائمة على الأدلة. وينظر إلى أساس هذه التقارير التجارية على أنها مصادرُ مختلفة للبيانات القادمة من داخل أو خارج المؤسسة (أنظمة معالجة العمليات التجارية عبر الإنترنت [OLTP]). ويتضمَّن إنشاءُ هذه التقارير إجراءات ETL (الاستخراج، والتحويل، والتنزيل) بالتنسيق مع مستودع البيانات، ثم استخدام واحدٍ أو أكثر من أدوات إعداد التقارير (انظر الفصل ٣ للحصول على وصفٍ مُفصَّل لهذه المفاهيم).

وبسبب التوسُّع السريع في تقنية المعلومات بجانب الحاجة إلى تحسين القدرة التنافسية في قطاع الأعمال؛ فقد حدثت زيادةٌ في استخدام القدرة الحاسوبية لإنتاج تقارير مُوحَّدة تنضمُّ إلى وجهات نظر مختلفة للمشروع في مكانٍ واحد. وعادةً ما تتضمَّن عملية إعداد التقارير هذه الاستعلام عن مصادر البيانات المهيكلة، والتي تمَّ إنشاء معظمها باستخدام نماذج بيانات منطقية وقواميس بيانات مختلفة؛ وذلك لإنتاج تقرير يسهل قراءته واستيعابه. كما تسمح هذه الأنواع من تقارير الأعمال للمديرين وزملاء العمل بالبقاء على دراية ومشاركة والقدرة على النظر في الخيارات والبدائل لاتخاذ قرارات صائبة. ويوضِّح الشكل ٢-١٨ دورةً مستمرةً من الحصول على البيانات ← توليد المعلومات ← صُنع القرار ← إدارة عمليات الأعمال. ولعل المهمة الأكثر حساسيةً في هذه العملية الدورية؛ هي إعداد التقارير (أيْ: توليد المعلومات) - وتحويل البيانات من مصادر مختلفة إلى معلومات قابلة للتطبيق.

إنَّ مفاتيح النجاح لأيِّ تقرير هي الوضوح والإيجاز والاكتمال والصواب. وتتغير طبيعة التقرير ومستوى أهمية هذه العوامل المؤدية إلى النجاح بشكلٍ كبير بناءً على «لمن يتم توجيه هذه

التقارير؟». يتمُّ توجيه معظم البحوث في التقارير الفعَّالة نحو التقارير الداخلية التي تمدُّ أصحاب المصلحـــة وصانعـــي القرار داخل المنظمـــة بالمعلومات. وهناك أيضًا تقاريـــر خارجية بين الشركات والحكومة (مثل: تقارير لأغراض الضريبة أو للإيداعات المنتظمة إلى لجنة الأوراق المالية والبورصات).



شكل ٢-١٨: دور إعداد تقارير المعلومات في اتخاذ القرارات الإدارية

وعلى الرغم من وجودِ تشكيلةٍ عريضةٍ من التقارير التجارية؛ فإنه يمكِن تصنيفُ تلك التي يتم استخدامها غالبًا لأغراض إدارية إلى ثلاث فئات رئيسة (Hill, 2016).

- **تقاريـــر إدارية موزونة:** في العديد من المؤسســـات؛ تتـــمُّ إدارة أداء الأعمال من خلال مقاييس مُوجَّهة باتجاه النتائج. بالنســـبة للمجموعات الخارجية؛ تكون هذه الاتفاقيات على مســـتوى الخدمة. أمَّا بالنســـبة للإدارة الداخلية؛ فتُعَدُّ مؤشرات الأداء الرئيسة (KPIs). وعادةً ما يكون هنـــاك أهدافٌ متفقٌ عليها على نِطاق المؤسســـة؛ ليتم تعقُّبها على مدى فترة زمنية ما. ويمكن استخدامها كجزءٍ من إستراتيجيات الإدارة الأخرى مثل Six Sigma أو إدارة الجودة الشاملة.
- **تقارير من نوع لوحة المعلومات:** كانت الفكرةُ الشائعة في التقارير التجارية في السنوات الأخيرة، هـــي تقديم مجموعةٍ من مـــؤشرات الأداء المختلفة على صفحة واحدة، مثـــل لوحة القيادة في الســـيارة. وعادةً ما يوفر منتجو لوحة المعلومات مجموعةً من التقارير المُحَدَّدة مسبقاً بعناصر ســـاكنة وهيكل ثابت، مع الســـماح أيضًا بتخصيص احتياجات لوحـــة المعلومات وطرق العرض ووضع أهداف لمقاييس مختلفة. ومن الشـــائع أن تكون هناك إشـــارات ضوئية مرورية ملونة كإشارات المرور المُحدَّدة للأداء (الأحمر، الأصفر، الأخضر)؛ وذلك لتوجيه انتباه الإدارة إلى مناطق معينة. وسيتمُّ الخوض بتفاصيل أكثر عن لوحات المعلومات في جزءٍ لاحق من هذا الفصل.

- **تقاريـــر متوازنـــة من نوع بطاقات الأداء:** قام بتطوير هـــذه الطريقة Kaplan and Norton، وهي طريقةٌ تحاول تقديم رؤية متكاملة للنجاح في أيِّ مؤسســـة. وإضافةً إلى الأداء المالي؛ فإنَّ التقاريـــر المتوازنة من نوع بطاقات الأداء تتضمَّـــن أيضًا العملاء وعملية الأعمال ووجهات نظر للتعلُّم والنمو. وسنعرض بمزيدٍ من التفصيل بطاقات الأداء المتوازنة لاحقًا في هذا الفصل.

تقـــدِّم الحالـــة العملية ٢-٥ مثالًا لتوضيح قوة وفائدة إنشـــاء التقاريـــر الآلية (وفي وقتٍ من الأزمات الطبيعية والفوضوية إلى حَدٍّ ما) لمنظمة كبيرة، مثل: FEMA.

حالة عملية ٢-٥

انتهاء فيضان الورق في FEMA

حصل الموظفون في وكالة إدارة الطوارئ الفيدرالية (FEMA) - وهي وكالةٌ فدرالية أمريكية تنسِّـــق الاســـتجابة للكوارث عندما يعلن الرئيسُ عن وقوع كارثة وطنية - على فيضانين في آنٍ واحد. أولًا: غطَّت المياه الأرض. بعد ذلك؛ غطَّى عددٌ كبيرٌ من الأوراق المطلوبـــة من جانـــب إدارة البرنامج الوطني للتأمين ضد الفيضانـــات (NFIP) مكاتب هـــؤلاء الموظفين؛ إذ تمَّ سَـــكْبُ حزم كثيرة مـــن التقارير المخططـــة باللون الأخضر من الطابعة المركزية في تلك المكاتب. كانت التقارير الفردية في بعض الأحيان ذات سُـــمْك يبلـــغ ١٨ بوصة، مع مجموعـــة صلبة من المعلومات حول مطالبات التأمين أو أقســـاط التأمين، أو مدفوعات مدفونة فيها في مكانٍ ما.

لا يدَّعـــي بيـــل بارتون ومايك مايلز أنهما قادران على فعـــل أيِّ شيء حيال الطقس؛ ولكن قام مديرُ المشروع وعالمُ الحاسب، على التوالي، من مؤسسة علوم الحاسب (CSC) باســـتخدام برنامج WebFOCUS مـــن "بناة المعلومات" لإعـــادة تدفق فيضان الورق بواســـطة NFIP. يســـمح البرنامجُ للحكومة بالعمل مع شركات التأمين الوطنية؛ لجمع أقســـاط التأمين ضـــد الفيضانات ودفع المطالبات للفيضانـــات في المجتمعات التي تبني تدابير معينة للتحكم في الفيضانات. وكنتيجة لعمل لجنة CSC؛ لم يَعُد موظفو FEMA بحاجـــةٍ إلى تصفـــح تقارير ورقية للعثور على البيانات التـــي يحتاجون إليها. وبدلًا من ذلك؛ يتصفحون بيانات التأمين المنشـــورة عـــلى موقع NFIP's BureauNet، ويختارون فقط المعلومات التي يرغبون في رؤيتها، ويحصلون منها على تقريرٍ معروض على شاشة أو يقومون بتنزيل البيانات في صورة جدول. وهذه مجرد بداية للمدخرات التي قدَّمها WebFOCUS. وقـــد انخفض عددُ المرات التـــي يطلب فيها موظفو NFIP من CSC

لتقارير خاصة إلى النصف؛ لأن موظفي NFIP صار بإمكانهم إنشاءُ العديد من التقارير الخاصة التي يحتاجون إليها دون استدعاء مبرمج لتطويرها. إضافةً إلى تكلفة إنشاء BureauNet في المقام الأول. ويُقدِّر بارتون أن استخدام برمجيات الويب وقاعدة البيانات التقليدية لتصدير البيانات من الإطار المركزي لـ FEMA، وتخزينها في قاعدة بيانات جديدة، وربطها بخادم ويب - قد يُكلِّف قرابة الـ ١٠٠ ضعف - أي أكثر من ٥٠٠٠٠٠ دولار - ويحتاج إلى حوالي عامين لإكماله، مقارنةً بالأشهر القليلة التي قضاها مايلز للوصول إلى الحل عن طريق WebFOCUS.

عندما هبَّت عاصفة أليسون الاستوائية، وهي عبارة عن مجموعة سحابية ضخمة من الغيوم الدوَّارة، وتحرّكت من خليج المكسيك على ساحل تكساس ولويزيانا في يونيو ٢٠٠١، قتلت ٣٤ شخصًا معظمهم من الغرق، كما دمَّرت وحطَّمَت ١٦٠٠٠ منزل وشركة، كما شرَّدت أكثر من ١٠٠٠٠ عائلة. وأعلن الرئيس جورج دبليو بوش مقاطعة تكساس ٢٨ كمنطقة كوارث، وتحرَّكت FEMA للمساعدة. كان هذا أول اختبار جاد لـ BureauNet، وتم تسليمه بالفعل. وقد أدَّى هذا الاستخدام الشامل الأول لـ BureauNet إلى وصول موظفي FEMA الميدانيين بسهولة إلى ما يحتاجون إليه عند الحاجة إليه، وطلب العديد من أنواع التقارير الجديدة. ولحُسْن الحظ، كان مايلز وWebFOCUS على مستوى المهمة. يقول بارتون: «في بعض الحالات سوف تطلب FEMA نوعًا جديدًا من التقارير يومًا ما، وسيحصل عليه مايلز على BureauNet في اليوم التالي؛ وذلك بفضل السرعة التي يمكنه بها إنشاء تقارير جديدة في WebFOCUS».

وأشار بارتون إلى أن الطلب المفاجئ على النظام لم يكن لـه تأثيرٌ كبيرٌ على أدائه. ويضيف: «لقد تعاملنا مع الطلب على ما يُرام؛ فلم تكن لدينا مشكلات في ذلك على الإطلاق. وقد أحدثنا فرقًا هائلًا لـ FEMA وللوظيفة التي كان عليهم القيامُ بها. إذ لم يكن لديهم هذا المستوى من الوصول من قبل، وأيضًا لم يكن بإمكانهم إنشاء مثل هذه التقارير المفصَّلة والمحدَّدة بمجرد النقر بالماوس على سطح مكتبهم كما هو الحال الآن».

أسئلة للمناقشة:

١- ما هي FEMA، وماذاتفعل؟

٢- ما هي التحديات الرئيسية التي تواجهها FEMA؟

٣- كيف قامت FEMA بتحسين ممارساتها غير الفعَّالة لإعداد التقارير؟

Source: Information Builders success story. Useful information flows at disaster response agency. informationbuilders.com/applications/fema (accessed May 2016); and fema.gov.

أسئلة مراجعة على القسم ٢-٧:

١- ما هو التقرير؟ وفيمَ تُستَخدَم التقارير؟

٢- ما هو تقرير الأعمال؟ وما هي الخصائص الرئيسة لتقرير الأعمال الجيد؟

٣- قم بوصف العملية الدورية للإدارة. مع التعليق على دور تقارير الأعمال.

٤- اذكر مع الشرح الفئات الثلاث الرئيسة لتقارير الأعمال.

٥- ما هي المكوِّنات الرئيسة لنظام إعداد تقارير الأعمال؟

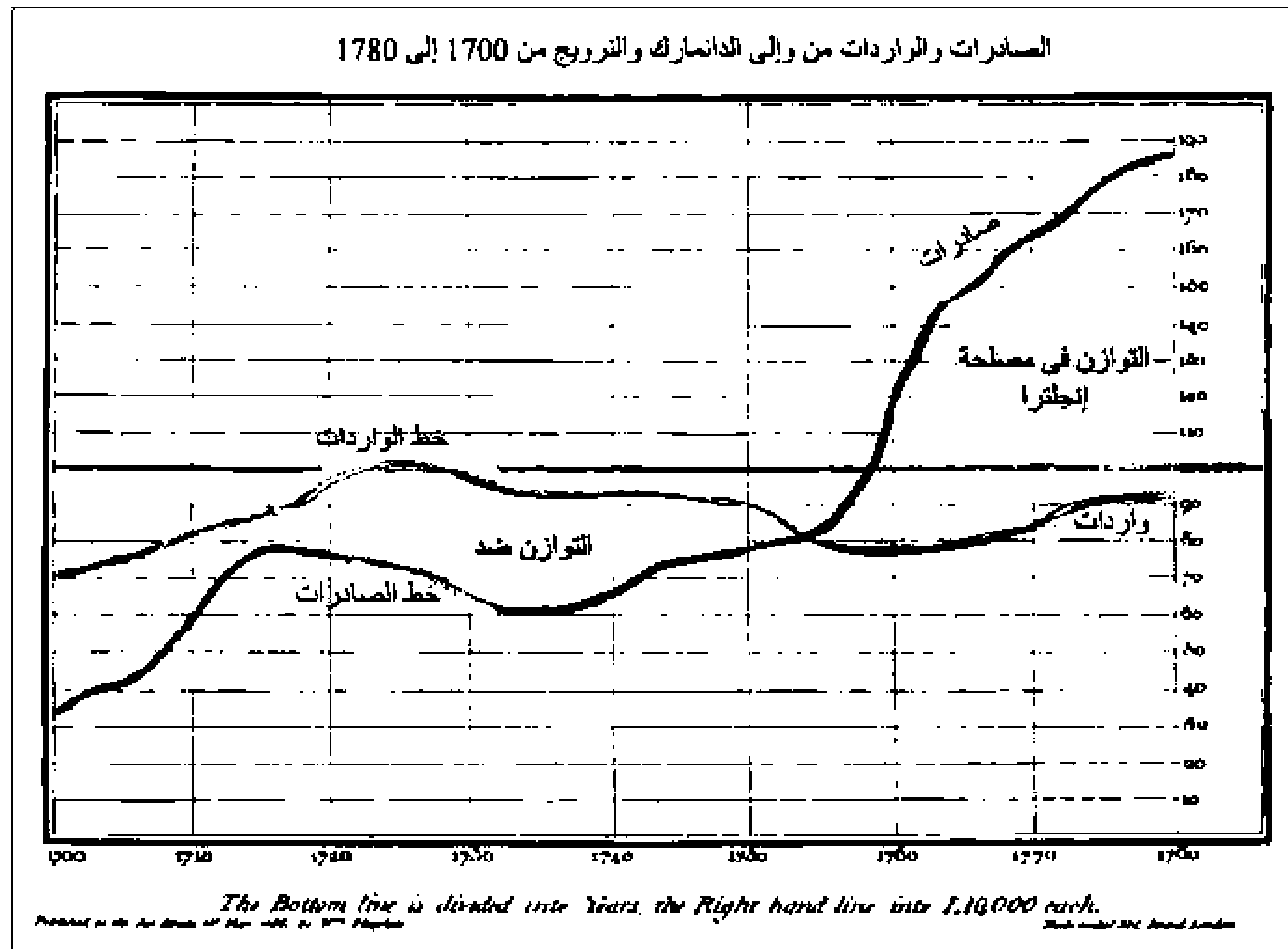
٢-٨ تصوير البيانات:

تمَّ تعريف تصوير البيانات (أو بشكلٍ أكثر ملاءمةً، التصوير المعلوماتي) على أنه «استخدام التمثيلات المرئية لاستكشاف، ووضع منطق، ونقل البيانات» (Few, 2007). وعلى الرغم من أن الاسم المُستَخدَم بشكلٍ شائع هو تمثيل البيانات؛ فإن المقصود من هذا هو تصوير المعلومات. ولأن المعلومات هي تجميع وتلخيص ووَضْع البيانات في سياقها (حقائق أولية)؛ فإن ما يتمُّ تصويره هو المعلومات وليس البيانات. ومع ذلك؛ ونظرًا لاستخدام المصطلحين تصوير البيانات وتصوير المعلومات بشكلٍ متبادل ومرادف؛ فإننا سنتبع ذلك في هذا الفصل.

يرتبط تصويرُ البيانات ارتباطًا وثيقًا بمجالات الرسومات البيانية للمعلومات، وتصوير المعلومات والتصوُّر العلمي والرسومات البيانية الإحصائية. وحتى وقتٍ قريب؛ كانت الأشكال الرئيسة لتصوير البيانات المتاحة في كلٍّ من تطبيقات ذكاء الأعمال تتضمَّن الجداول والرسوم البيانية؛ إضافةً إلى الأنواع الأخرى من العناصر المرئية المُستَخدَمة لإنشاء بطاقات الأداء ولوحات البيانات. ومن أجل فهمٍ أفضل للاتجاهات الحالية والمستقبلية في مجال رؤية البيانات؛ فنرى هنا أنه من الأفضل البدء ببعض السياق التاريخي.

لمحة تاريخية عن تصوير البيانات:

عــلى الرغم من حقيقة أن تاريخ تصوير البيانات يعود إلى القــرن الثاني الميلادي؛ فإنَّ معظم التطــوُّرات التي حدثت في المئتين وخمســين عامًا الماضية؛ حدثت في الغالب خلال الثلاثين ســنةً الأخــيرة فقــط (Few, 2007). وعلى الرغم من أن التصوير المرئي لم يُعترف به على نطاقٍ واســعٍ كنظام حتى وقت قريب؛ فإنَّ الأشــكال المرئية الأكثر شــيوعًا اليوم تعود إلى بضعة قرون ســابقة. وقد شجَّع الاستكشافُ الجغرافي، والرياضيات، والتاريخ الشعبي على إنشاء خرائط مبكرة، ورسوم بيانية، وجداول زمنية تعود إلى القرن الســابع عشر؛ غير أنَّ الفضل يُنسَــب إلى ويليام بلايفير على نطاقٍ واسعٍ كمخترع الرسم البياني الحديث؛ وذلك بعد أن أنشأ أول شريط رسومات بيانية مُوزَّع على نطاق واســع في أطلسه التجاري والسياسي لعام ١٧٨٦م، وهو ما يُعَد بشكلٍ عام أولَ سلسلة زمنية تصوِّر الرسومات البيانية في كتابه الإحصائي، الذي نُشِرَ في عام ١٨٠١م (انظر: الشكل ٢-١٩).



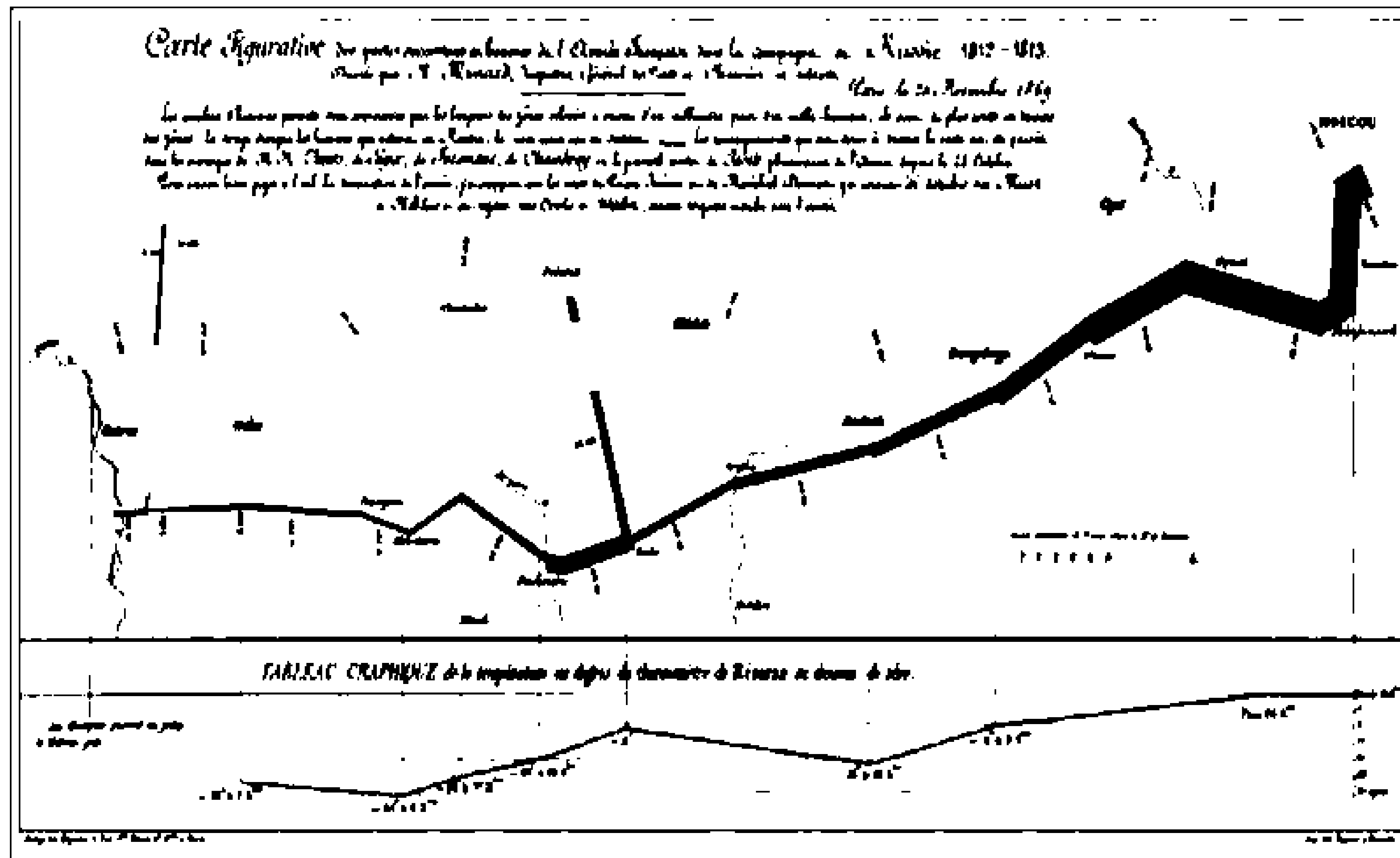
شكل ٢-١٩: خط بياني لأول سلسلة زمنية عن طريق WILLIAM PLAYFAIR في ١٨٠١م

ولعــل أبرز مبتكرٍ للرســومات البيانيــة للمعلومات خلال هذه الفترة؛ كان تشــارلز جوزيف مينــارد، الذي صوَّر بيانيًّا الخســائر التي تكبَّدها جيش نابليون في الحملة الروســية عام ١٨١٢م (انظر: الشكل ٢-٢٠). بدايةً من الحدود البولندية - الروسية، يوضِّح الخط السميك حجم الجيش

في كلِّ موقع. ويظهر مسار تراجع نابليون من موسكو في فصل الشتاء القارس من خلال الشريط السُّفلي المظلم، المرتبط بدرجات الحرارة والفترات الزمنية. يقول خبير التصوير البياني والمؤلف والناقد الشهير إدوارد توفت أنَّ هذا «قد يكون أفضل رسمٍ بياني إحصائي يتمُّ رسمه على الإطلاق». في هذا الرسم البياني؛ تمَّ تمثيل Minard آنيًّا بعِدَّة أبعاد للبيانات (مثل: حجم الجيش، واتجاه الحركة، والمواقع الجغرافية، ودرجة الحرارة الخارجية، وما إلى ذلك) بطريقة فنية وإعلامية. وقد شهد القرن العشرون بزوغ سلوك جديد للتصوير البياني يتسمُ بأنه أكثر رسميةً وتخطيطًا، والذي كان يميلُ إلى التركيز على جوانب، مثل: اللون، ومقاييس القيمة، ووضع العلامات. وفي منتصف القرن العشرين، أصدر رسام الخرائط والباحث النظري جاك بيرتين كتابه «علم الرسم البياني»، والذي يقول عنه البعض: إنه يُعتبَر بمثابة الأساس النظري للتصوير الحديث للمعلومات. وعلى الرغم من أن معظم أنماطه قد تقادمت بسبب الأبحاث الحديثة أو عدم قابليتها للتطبيق بشكلٍ كامل على الوسائط الرقمية؛ فإنَّ الكثير منها ما زال مناسبًا للغاية.

ومع بدايات الألفية الجديدة؛ ظهر الإنترنت كوسيطٍ جديدٍ للتصوير البياني، وجَلَبَ معه الكثير من الحِيل والقدرات الجديدة. ولم يجعل التوزيع الرقمي كلًّا من البيانات والتصوير البياني على مستوى العالم أكثر سهولةً للوصول إلى جمهور أوسع (أيْ: زيادة محو الأمية المرئية)؛ بل حفَّز أيضًا تصميم نماذج جديدة تتضمَّن التفاعل، والرسوم المتحركة، وعرض الرسومات البيانية للتقنية الفريدة من نوعها لشاشات الوسائط، وتغذية البيانات بشكلٍ فوري لإنشاء بيئات غامرة للتواصل واستهلاك البيانات.

وقد اهتمَّت الشركاتُ والأفراد بالبيانات بشكلٍ مفاجئ، وهذا الاهتمام أثارَ بدوره الحاجة إلى أدوات مرئية تساعدهم على فهمها. وتعمل أجهزة الاستشعار الرخيصة الثمن مع فلسفة «قم ببناء نظامك الخاص بنفسك» على تقليل تكاليف جمع البيانات ومعالجتها. وقد أدَّى ظهور عددٍ لا يُحصَى من التطبيقات الأخرى، وأدوات البرامج، ومكتبات الشفرة منخفضة المستوى إلى مساعدة الناس في جمع البيانات وتنظيمها ومعالجتها وتصويرها وفهمها عمليًّا من أيِّ مصدر. وقد كان الإنترنت أيضًا بمثابة قناة توزيع رائعة للتصويرات البيانية؛ فقد تمَّ تجميع مجموعة متنوعة من المُصمِّمين والمبرمجين ورسامي الخرائط والمشتغلين بالصيانة، والبيانات المفككة لنشر كلِّ أنواع الأفكار والأدوات الجديدة للعمل مع البيانات في كلٍّ من الأشكال المرئية وغير المرئية.



شكل ٢-٢٠: هلاك جيش نابليون في أثناء الحملة الروسية لعام ١٨١٢م

كما قامت خرائطُ Google أيضًا بمشـاركة فرديــة في كلٍّ من اتفاقيات الواجهة (انقر للتنقل، انقــر نقرًا مزدوجًا للتكبير) والتقنية (خريطة بحجم ٢٥٦ بكســل مربع مع أســماء ملفات يمكن التنبؤ بها) لعرض الجغرافيا التفاعلية عبر الإنترنت، بالشــكل الذي يعلم به معظم الناس ما يجب عليهم فعله عندما يتمُّ تقديم خريطة عبر الإنترنت. لقد قدَّم Flash خدمةً جيدةً كمنصة تصفُّح لتصميم وتطويــر تطبيقات إنترنت غنية وجميلة تتضمَّن تصويرات وخرائط بيانات تفاعلية؛ وفي الوقــت الحالي؛ بدأت تقنيات جديدة للمتصفحات في الظهور، مثل: Canvas، وSVG (والتي يتمُّ تضمينها مجتمعةً في بعض الأحيان تحت مظلة HTML5) لتحَدِّي السيادة التي حظي بها فلاش والعمل على زيادة فرص وصول واجهات التصوير المتحرك إلى أجهزة التليفون المحمولة.

يصعب التنبؤ بمســتقبل تصوير البيانات/ المعلومات. يمكننا فقط أن نستنتج مما تمَّ اختراعه بالفعل: المزيد من التصوير ثلاثي الأبعاد، وتجربة مليئة بالبيانات المتعددة الأبعاد في بيئة الواقع الافــتراضي، والتصوير البياني المُجسَّــم للمعلومات. وهناك فرصة جيدة بأن نرى شــيئًا جديدًا لم نشهده من قبل في عالم تصوير المعلومات الذي تم اختراعه قبل نهاية هذا العقد. وتوضِّح الحالة العمليــة ٢-٦ كيــف أن أدوات التحليلات/ التقاريــر المرئية مثل Tableau يمكنها أن تســاعد في تسهيل اتخاذ القرار بفعالية وكفاءة من خلال إنشاء ومشاركة المعلومات/ المعرفة.

حالة عملية ٢-٦

ماكفارلان سميث يحسن رؤية الأداء التشغيلي باستخدام Tableau Online

معلومات أساسية:

حصل مكفارلان سـميث على مكانةٍ في التاريخ الطبي؛ إذ عقدت الشركة تعيينًا ملكيًا لتوفير الدواء لجلالة الملكة فيكتوريا وزوّدت أخصائي التوليد الرائد السير جيمس سيمبسون بالكلوروفـورم لتجاربه في تخفيف الألم في أثناء المخاض والـولادة. وتُعَدُّ اليوم ماكفارلان سـميث شركة تابعـة لشركـة Fine Chemical وCatalysts التابعـة لشركة Johnson Matthey plc. وتُعَـدُّ الشركـة المصنِّعة للأدوية؛ هي الشركـة الرائدة في العالم في تصنيع المخدِّر الأفيوني، مثل: الكودايين والمورفين. ويقوم موظفو مكفارلان سميث كلَّ يوم باتخاذ قرارات بنـاءً على بياناتهم. فهم يقومـون بجمع وتحليل البيانات التشـغيلية للتصنيع، مثل بيانات تسـمح لهم بتحقيق أهداف التحسـين المسـتمر. وتعتمد كلٌّ من المبيعات والتسـويق والتمويل على البيانـات؛ لتحديد فرص الأعمال التجاريـة الدوائية الجديدة، وتنمية العائدات الاسـتثمارية وتلبية احتياجات العمـلاء. وإضافةً إلى ذلك؛ يحتاجُ فرع الشركة في إدنبره إلى مراقبة بيانات الجودة واتجاهها، وإعداد تقارير عنها؛ لضمان هويتها وجودتهـا ونقـاء كلِّ مكوناتها الدوائية للعملاء والسُّـلطات التنظيمية، مثل إدارة الغذاء والدواء الأمريكية (FDA) وغيرها كجزءٍ من ممارسات التصنيع الجيدة (cGMP).

التحديات: مصادر متعددة للحقيقة والبطء، وعمليات إعداد التقارير المرهقة:

لم تكن عملية جمع تلك البيانات واتخاذ القرارات، وإعداد التقارير سهلةً على الرغم من ذلك. فقد كانت البيانات الخاصة بالنشـاط التجـاري مبعثرةً، بما في ذلك توصيات الشركة بشأن تخطيط موارد المشروع (ERP)، داخل قواعد البيانات الإدارية القديمة مثل SQL، وقواعد بيانات Access، وجداول البيانات المسـتقلة. وعندما تمَّ الاحتياجُ لهذه البيانات لصُنع القرار، تمَّ تكريس وقت طويل جدًا وتخصيص موارد مفرطة للتنقيب في البيانات ودمجها وعرضها في جدول بيانات أو في أيِّ وسيلة عرض أخرى.

وكانـت جودة البيانات مصدر قلقٍ آخر. ونظرًا لاعتماد فِرق العمل على مصادرهم الفرديـة للبيانات؛ فقد كانـت هناك إصدارات مُتعدِّدة مـن الحقيقة والتعارضات بين البيانـات. وكان من الصعب أحيانًـا معرفة أيِّ إصدار من البيانات كان صحيحًا وأيها لم يكن كذلك.

لم يتوقــف الأمــرُ عند هذا الحد؛ فحتــى بعد جمع البيانات وعرضهــا، كان إجراء تغييرات «على الطاير» بطيئًا وصعبًا. وفي الواقع؛ كلما حاول أحد أعضاء فريق ماكفارلان ســميث عمل اتجاه للبيانات أو إجــراء أيِّ تحليل آخر؛ فإن التغيــيرات التي تطرأ على البيانات كانت تحتاج إلى الموافقة عليها. وكانت النتيجة النهائية هي أن البيانات كانت متقادمةً بشكلٍ مؤثر جدًا في الوقت الذي كانت تستخدم في صُنْع القرار.

ويُســلِّط ليام ميلز، رئيس قسم التحســين المستمر في ماكفارلان سميث، الضوءَ على ســيناريو نموذجي لإعداد التقارير؛ فيقول: «تتمثل إحدى عملياتنا الرئيســية في إعداد التقاريــر في الإجــراء التصحيحي والإجــراء الوقائي، أو CAPA، وهو عبــارة عن تحليلٍ لعمليــات التصنيع في ماكفارلان ســميث التي تمَّ اتخاذها لإزالة أســباب عدم التوافق أو حالات أخرى غير مرغوب فيها. كما تمَّ تخصيصُ مئات الســاعات كل شــهر لسحب البيانــات آنيًا من أجل CAPA - واســتغرق الأمر أيامًا لإنتــاج كل تقرير. وكان تحليلُ الاتجاهــات صعبًــا أيضًا؛ لأن البيانات كانت ثابتة (ســاكنة). وفي ســيناريوهات أخرى لإعداد التقارير، غالبًا ما كان يتحتم علينا انتظارُ تحليل جدول محوري لجدول البيانات؛ والذي تمَّ تقديمه بعد ذلك في صورة رســم بياني، كما تمَّت طباعته، وتثبيته على الحائط؛ كي يراجعه الجميع بسهولة».

كانت عمليات إعداد التقارير البطيئة والمكثفة للعمالة، والنسخ المختلفة للحقيقة، والبيانات الســاكنة، كانت كلها عوامــل محفزة للتغيير. يقول ميلز: «شــعر الكثيرون بالإحبــاط؛ لأنهم اعتقدوا أنهم ليس لديهــم صورة كاملة عن العمل. لقد أجرينا الكثير والكثــير من المناقشــات حول القضايــا التي واجهناها - عندمــا كان علينا التحدُّث عن تقارير ذكاء الأعمال».

الحل: تصويرات البيانات التفاعلية:

كان لدى أحد أعضاء فريق ماكفارلان ســميث خبرة ســابقة في استخدام Tableau، وأيضًا استكشــاف الحلول المُوصَى بها من قِبل ميلز. وسرعان ما أقنعت نسخة تجريبية مجانيــة من Tableau Online ميلز بأنَّ حلَّ إضافــة التصوير التفاعلي للبيانات يمكنه التغلُّب على مشكلات البيانات التي كانوا يواجهونها.

ويضيــف: «لقــد فزت في الحال تقريبًا. فإن ســهولة الاســتخدام والأداء الوظيفي واتســاع تصويرات البيانات كلها مثــيرة جدًا للإعجاب. وبالطبــع؛ كونها حلاً قائماً على

البرمجيات كخدمة (SaaS)؛ فإنه لا يُوجَد أيُّ استثمار في البنية التحتية التقنية يمكننا أن نعيشه على الفور تقريبًا، وأيضًا لدينا المرونة لإضافة المستخدمين كلما احتجنا إلى ذلك».

إن أحد الأسـئلة الرئيسـية التي يجبُ الإجابة عنها تتعلق بأمان البيانات عبر الإنترنت. «لدى الشركة الأم جونسـون ماثي إستراتيجية السـحابة؛ ولكن يجب أن نكون متأكدين من أن إضافـة أيِّ حلٍّ آمن تمامًا. إن ميزات Tableau Online مثل تسـجيل الدخول الفردي والسَّماح للمستخدمين المصرَّح لهم فقط بالتعامل مع البيانات توفر الأمان والثقة اللازمين».

أمَّا سؤال الأمان الآخر الذي كان ماكفارلان سميث وجونسون ماثي يريدان الإجابة عنه: أين يتـمُّ تخزين البيانات ماديًّا؟ يضيف ميلز: «نحن راضون عن أداء Tableau Online، والذي يسـتوفي المعايير الخاصة بنا لأمان البيانات والخصوصية؛ إذ تتم اسـتضافة البيانات والمصنَّفات في مركز بيانات دبلن الجديد في Tableau، لذلك فهي لا تغادر أوروبا أبدًا».

وبعد تجربةٍ دامت سـتة أسـابيع، عمل خلالها مدير المبيعـات في **Tableau** مع ميلز وفريقـه لبناء قضية تجارية لـ **Tableau Online**، وافق فريق الإدارة على ذلك على الفور، وبدأ برنامج تجريبي يضمُّ ١٠ مستخدمين. وشملت التجربة مبادرة لتحسين جودة التصنيع: بالنظر إلى الانحرافات عن القاعدة، كأن يتجاوز جهاز التسـخين المستخدَم في عملية تصنيع المخدرات الأفيونية درجة الحرارة. من هذا المنطلق؛ تمَّ إنشـاء لوحة معلومات «عمليات الجودة» لتتبُّع وقياس الانحرافات ووضع إجراءات لتحسين الجودة والأداء التشغيلي.

يقول ميلز: «أشارت لوحة المعلومات هذه على الفور إلى أماكن الانحرافات. لم نكن ننشط عبر صفوفٍ من البيانات - لقد توصَّلنا إلى إجابات فورية».

وخلال هذه التجربة الأوَّلية والتجريبية؛ استخدم الفريق أدوات التدريب Tableau، مثل مقاطـع الفيديو التدريبية المجانية، وإرشـادات المنتجـات والتدريب المباشر عبر الإنترنت. كما شـاركوا في حدث «التدريب الأسـاسي» على مدى يومين في لندن. ووفقاً لميلـز: «كان التدريب دقيقًا وموجهًا فقط على المسـتوى الصحيح. وقد أظهر للجميع فقـط مدى سـهولة Tableau Online. فقـد أصبح بإمكاننا تصوير ١٠ سـنوات من البيانات في عددٍ قليل من النقرات». لدى الشركة الآن خمسة مستخدمين لسطح مكتب Tableau، وما يصل إلى ٢٠٠ مستخدم مرخص لـ Tableau Online.

وبشكلٍ خاصٍّ يشبه ميلز وفريقه ملامح Tableau Union في الإصدار ٩,٣؛ مما يتيح لهم ضمَّ البيانات التي تمَّ تقسـيمها إلى ملفات صغيرة. ويضيـف: «من الصعب أحيانًا تجميع

البيانات التي نستخدمها للتحليل؛ إذ تتيحُ لنا مَيْزة الاتحاد العمل مع البيانات المنتشرة عبر عِدة علامات تبويب أو ملفات؛ مما يقلل من الوقت الذي نقضيه في إعداد البيانات.

النتائج: تحليلات السَّحابة تقوم بتحويل صُنع القرار وإعداد التقارير:

من خلال توحيد المعايير على Tableau Online، حول ماكفارلان سميث سرعة ودقة اتخاذ القرار وإعداد تقارير الأعمال. وهذا يتضمَّن:

- يمكن إنتاجُ لوحات تفاعلية جديدة خلال ساعة واحدة. في السابق؛ كان يستغرق الأمرُ أيامًا لدمج البيانات وعرضها في جدول بيانات ثابت.
- يمكن الآن إنتاجُ تقرير عملية تصنيع CAPA في دقائق، والذي كان يستغرق إنتاجه في السابق المئات من ساعات العمل كل شهر - مع الأفكار المشتركة في السحابة.
- يمكِن تغيير التقارير والاستعلام «على الطاير» بسرعة وسهولة، دون تدخل تقني؛ إذ يتمتع ماكفارلان سميث بالمرونة لنشر لوحات المعلومات مع Tableau، ومشاركتها مع الزملاء أو الشركاء أو العملاء.
- تمتلك الشركة نسخةً واحدةً وموثوقةً من الحقيقة.
- تجري ماكفارلان سميث حاليًّا مناقشات حول بياناتها - وليس حول القضايا المتعلقة بتكامل وجودة البيانات.
- يمكِن جلب المستخدمين الجدد عبر الإنترنت على الفور تقريبًا - وليس هناك بنية أساسية تقنية لإدارتها.

وبعد هذا النجاح المبدئي؛ يقوم ماكفارلان سميث الآن بتوسيع نطاق Tableau Online إلى إعداد التقارير المالية، وتحليلات سلاسل التوريد وتوقعات المبيعات. ويختتم ميلز قائلًا: «تستند إستراتيجية أعمالنا الآن على قرارات تعتمد على البيانات وليس على الآراء. كما تُمكّننا التصويرات التفاعلية من تحديد الاتجاهات بشكلٍ فوري، وتحديد التحسينات في العملية واستخلاص ذكاء الأعمال إلى المستوى التالي. أستطيع الآن تعريف مهنتي بواسطة Tableau».

أسئلة للمناقشة:

١- ما هي التحديات المتعلقة بالبيانات وإعداد التقارير التي واجهت ماكفارلان سميث Macfarlan Smith؟

٢- ما هو الحل وماذا كانت النتائج/ الفوائد التي تمَّ الحصول عليها؟

Source: Tableau Customer Case Study, "Macfarlan Smith improves operational performance insight with Tableau Online," http://www.tableau.com/stories/customer/macfarlan-smith-improves-operational-performance-insight-tableau-online (accessed October 2016).

أسئلة مراجعة على القسم ٢-٨:

١- ما هو تصوير البيانات؟ ولماذا يتمُّ الاحتياج إليه؟

٢- ما هي الجذورُ التاريخية لتصوير البيانات؟

٣- حلِّل بعناية تصوير تشـارلز جوزيف مينارد الرسمى لمسيرة نابليون. مع تحديد والتعليق على جميع أبعاد المعلومات التي تمَّ التقاطها في هذا الرسم البياني القديم.

٤- مَنْ هو إدوارد توفت؟ وبرأيك لماذا علينا أن نتعرف على ما قام به من عمل؟

٥- ما هو في رأيك «الشيء الكبير التالي» في تصوير البيانات؟

٢-٩ الأنواع المختلفة للمخططات والرسوم البيانية:

غالباً مايكون المسـتخدمون النهائيون لأنظمة تحليلات الأعمال غير متأكدين من نوع الرسـم البيـاني أو الجـدول البياني الـذي يجب أن يسـتخدموه لغرض معين؛ إذ تكون بعض الرسـوم أو الجـداول البيانية هـي أفضل في الإجابة عن أنواع معينة من الأسـئلة؛ فقد يبـدو بعضها أفضل من غيرها ويبدو البعض الآخر بسـيطًا، ويكون البعض الآخـر معقدًا ومزدحمًا إلى حَدٍّ ما. وفيما يلي وصفٌ موجزٌ لأنواع الرسـوم البيانية و/ أو الجداول البيانية الشـائعة في معظم أدوات تحليل الأعمال وأنواع الأسئلة التي هي أفضل في الإجابة/ التحليل. وقد تمَّ تجميعُ هذه المواد من العديد من المقالات المنشورة وغيرها من المنشورات (أبيلا ٢٠٠٨؛ هاردن ٢٠١٢؛ ساس ٢٠١٤؛ وآخرون).

المخططات والرسوم البيانية الأساسية:

فيما يلي المخططات والرسومات البيانية الأساسية التي تُستخدَم عادةً لتصوير المعلومات.

خط الرسـم البياني (Line Chart): هو الصور البيانية الأكثر اسـتخدامًا لبيانات السلاسـل الزمنية. وتوضِّح المخططات الخطية (أو الرسوم البيانية الخطية) العلاقة بين متغيرين؛ وغالبًا ما يتمُّ

استخدامها لتتبُّع التغييرات أو الاتجاهات عبر الوقت (تعيين أحد المتغيرات إلى الوقت على المحور السـيني). وتعمل المخططات الخطية على توصيل نقاط بيانات فردية بشكلٍ متعاقب للمساعدة في استكشاف الاتجاهات المتغيرة على مدار فترة زمنية. وغالبًا ما يتمُّ استخدام المخططات الخطية؛ لإظهار التغييرات المعتمدة على الوقت في قِيَم بعض المقاييس، مثل: التغييرات على سعر سهم معين على مدار ٥ سنوات أو تغييرات على عددِ المكالمات اليومية لخدمة العملاء على مدار الشهر.

شريط الرسـم البياني (Bar Chart): تُعَدُّ من أكثر العناصر الأساسـية الأكثر اسـتخدامًا لتمثيل البيانات. تكون الرسـوم البيانية الشريطية فعَّالة عندما يكون لديك بيانات اسـمية أو بيانات رقمية تنقسم بشـكلٍ جيدٍ إلى فئات مختلفة؛ بحيث يمكنك مشاهدة النتائج والاتجاهات النسبية في بياناتك بسرعة. وغالبًا ما يتمُّ اسـتخدام المخططات الشريطية لمقارنة البيانات عبر فئات متعددة مثل النسبة المئوية للإنفاق الإعلاني بحسب الأقسام أو بحسب فئات المنتجات. يمكن أن تكون المخططات الشريطية عموديةً أو أفقية. ويمكن أيضًا تحزيمها بعضها فوق بعض؛ لإظهار أبعاد متعددة في مخططٍ واحد.

الرسـم البياني الدائري (Pie Chart): وكما يوحي اسـمها؛ فهي تكون جذَّابة للعين، مخطط دائري الهيئة. ولأنها جذابة جدًّا؛ فغالبًا ما تُسـتَخدَم بشـكلٍ غير صحيح. فعلى سبيل المثال: يمكن اسـتخدامها لعرض النسـبة المئوية النسـبية لميزانية الإعلان التي يتمُّ إنفاقهـا على خطوط الإنتاج المختلفة، أو يمكن أن تظهر النِّسَـب النسـبية من التخصصات التي التحق بها طلاب الجامعات في سنتهم الدراسية الثانية. فإذا كان عدد الفئات المراد إظهارها كبيراً إلى حَدٍّ ما (على سبيل المثال: أكثر من أربعة)؛ فيجب التفكير بجدية في استخدام شريط الرسم البياني بدلًا من الرسم البياني الدائري.

مربع التبعثر (Scatter Plot): غالبًا ما يتمُّ اسـتخدامها لاستكشـاف العلاقة بين متغيرين أو ثلاثة (في الصور ثنائية الأبعاد). ولأنها أدوات استكشـاف بصرية، ولديها أكثر من ثلاثة متغيرات؛ فإن ترجمتها إلى أكثر من ثلاثة أبعاد لا يمكن تحقيقها بسـهولة. وهي وسـيلةٌ فعَّالةٌ لاستكشاف وجود الاتجاهات والتركيزات والقِيم المتطرفة. فعلى سـبيل المثال: في الرسم البياني ذي المتغيرين (المحورين)، يمكن استخدام مربع التبعثر لتوضيح علاقة الارتباط بين عُمْر ووَزْن مرضى القلب، أو يمكـن أن يوضِّح العلاقة بين عـدد ممثلي خدمة العملاء وعدد مطالبات خدمة العملاء المفتوحة. وفي كثيرٍ من الأحيان؛ يتمُّ فرضُ خط اتجاه على مربع التبعثر ثنائي الأبعاد لتوضيح طبيعة العلاقة.

الرسـم البياني الفقاعي (Bubble Chart): هي غالبًا ما تكون عبارة عن نسخة مُحسَّنة من مربـع التبعـثر. وعلى الرغم من ذلك؛ فهي لا تُعَدُّ نمطًا مرئيًّا جديـدًا؛ بل ينبغي النظر إليها على أنهـا تقنيـةٌ لإثراء البيانات الموضَّحـة في مخططات التبعثر (أو حتى الخرائـط الجغرافية). ومن

خلال تفاوت حجم و/ أو لون الدوائر؛ يمكن إضافةُ أبعاد إضافية للبيانات؛ مما يعطي معنًى أكثر إثراءً حول البيانات. فعلى ســبيل المثال: يمكن اســتخدامُ رسم بياني فقاعي؛ لإظهار عرض تنافسي لحضور الفصل الدراسي على مستوى الكلية حسب التخصص، وحسب الوقت من اليوم، أو يمكن استخدامه لعرض هامش الربح حسب نوع المنتج وحسب المنطقة الجغرافية.

المخططات والرسوم البيانية المتخصصة:

إن الرســوم والمخططات البيانية التي نراجعها في هذا القســم؛ إما مســتمدة من المخططات الأساسية كحالات خاصة، أو أنها جديدةٌ نسبيًا، وهي خاصةٌ بنوع المشكلة أو منطقة التطبيق.

المــدرج التكراري (Histogram): من الناحية البيانية، يبــدو المدرج التكراري وكأنه شريط رســم بيــاني (bar chart). ويكمُن الفــرق بينهما فقط في المعلومات التي يتــمُّ تصويرها. ويتم استخدامُ المدرج التكراري S لإظهار التوزيع التكراري لمتغيرٍ أو لعدة متغيرات. في المدرج التكراري، غالبًا ما يُستَخدَم المحور السيني لإظهار الفئات أو النطاقات؛ في حين يتمُّ استخدامُ المحور الصادي لإظهار المقاييس/ القيم/ التكرارات. يُظهر المدرج التكراري S الشــكل التوزيعي للبيانات. وبهذه الطريقــة؛ من الممكــن إجراء فحص بصري لتوزيــع البيانات لمعرفة ما إذا كانــت موزعةً توزيعًا طبيعيًا أو أُسيًّا. فعلى سبيل المثال: يمكن استخدامُ المدرج التكراري لتوضيح أداء اختبار الفئة؛ إذ يمكِن عرضُ توزيع الدرجات إضافةً إلى التحليل المقارن للنتائج الفردية، أو يمكن استخدام المدرج التكراري لعرض توزيع أعمار قاعدة العملاء.

الرســم البياني جانت (Gantt): هي حالة خاصة من شرايط الرسومات البيانية الأفقية التي تُســتَخدَم لتصوير الجداول الزمنية للمشروع، ومدة مهام/ نشــاط المشروع، والتداخل بين المهام/ الأنشطة. ومن خلال عرض تواريخ/ أوقات البدء والانتهاء للمهام/ للأنشطة والعلاقات المتداخلة؛ تُقدِّم رسومات جانت مساعداتٍ لا تُقدَّر بثمن لإدارة المشاريع والتحكُّم فيها. فعلى سبيل المثال: غالبًا ما يتمُّ اســتخدامُ مخططــات جانت لعرض المخططات الزمنية للمــشروع وتداخلات المهامِّ وإكــمال المهامِّ النســبية (شريط جزئي يُوضِّح نســبة الإكمال داخل شريط يعــرض المدة الفعلية للمهمة) والموارد المخصصة لكلِّ مهمة والمعالم والإنجازات.

الرسم البياني بيرت (Pert): (وتُسمَّى أيضًا الرسوم البيانية للشبكة)، ويتمُّ تطويرها في المقام الأول؛ لتبســيط التخطيط، وجدولة المشــاريع الكبيرة والمعقدة. وهي تُظهر علاقات الأسبقية بين أنشــطة/ مهام المشروع. يتكوَّن الرســم بيرت من عُقَد (ممثلة بدوائر أو مســتطيلات) وحوافّ

(ممثلة بالأسـهم الموجهة). وبناءً على ما هو متعارف عليه بالنسـبة لبيرت؛ فإنه يمكن استخدامُ العقد أو الحوافّ لتمثيل أنشطة/ مهام المشروع (مخطط النشاط على أساس العُقَد مقابل النشاط على أساس التمثيل السهمي).

الخريطـة الجغرافية (Geographic map): عندما تتضمَّن مجموعـة البيانات أيَّ نوعٍ من بيانـات الموقع (مثل: العناوين المادية، أو الرموز البريدية، أو أسـماء الولايات، أو الاختصارات، أو أسـماء الدول، أو خطوط العرض/ الطول، أو بعض أنواع الترميـز الجغرافي المخصص)، ويكون من الأفضل والأكثر فائدةً الاطلاع على البيانات على خريطة. وعادةً ما يتم استخدام الخرائط بالتزامن مع الرسوم والمخططات البيانية الأخرى، بعضها في مقابل بعض. فعلى سبيل المثال: يمكن استخدامُ الخرائط لعرض توزيع طلبات خدمة العملاء على حسـب نوع المنتج (الموضَّح في الرسـوم البيانية الدائرية) على حسـب المواقع الجغرافية. وفي كثيرٍ مـن الأحيان؛ يمكن تصويرُ مجموعةٍ كبيرة من المعلومـات (مثل: التوزيـع العمري، أو توزيع الدخل، أو التعليم، أو النمو الاقتصادي، أو التغيرات السكانية) في خريطة جغرافية واحدة؛ للمساعدة في اتخاذ قرار بمكان فتح مطعم جديد، أو إنشاء محطة خدمة جديدة. وغالبًا ما تُسمَّى هذه الأنواع من النظم بنظم المعلومات الجغرافية (GIS).

مُخططـات بولت (Bullet): غالبًا ما يتمُّ اسـتخدام مخططـات Bullet؛ لإظهار التقدُّم نحو الهدف. وهي في الأساس شكلٌ من أشكال شريط الرسم البياني. وغالبًا ما يتمُّ استخدامها بدلًا من المقاييس، والعدادات، ومقاييس الحرارة في لوحة المعلومات لتوصيل المعنى المراد توصيله بشكلٍ أكثر بديهية ضمن مساحة أصغر بكثير. تقارن مخططات Bullet مقياسًا أساسيًّا (مثل: الإيرادات السنوية) مع واحدٍ أو أكثر من المقاييس الأخرى (مثل: هدف الإيرادات السنوي)، وتقدم ذلك في سـياق مقاييس أداء معينة (مثل: حصص المبيعات). ويمكن لمخططات Bullet أن توضِّح بشكلٍ بديهي كيفية أداء الإجراء الأساسي مقابل الأهداف العامة (مثل: مدى قرب مندوب المبيعات من تحقيق حصته السنوية).

الخريطة الحرارية (Heat Map): تُعَدُّ الخرائط الحرارية عناصر مرئية رائعة؛ لتوضيح المقارنة بين القيم المسـتمرة عبر فئتين باستخدام اللون. ويكون الهدف منها هو مساعدة المستخدِم على المشاهدة السريعة للمكان الذي يكون فيه تقاطع الفئات أقوى وأضعف من حيث القيم العددية للمقياس الجاري تحليله. فعلى سبيل المثال: من الممكن استخدامُ خرائط الحرارة؛ لإظهار تحليل تقسـيم الأسـواق المُسـتَهدَفة؛ بحيث يكون المقياس هو مبلغ الشراء (والذي يتم تمثيله بالتدرج اللوني)، وتكون الأبعاد هي العمر وتوزيع الدخل.

جداول التظليـــل (Highlight Table): وتهدف جداول التظليـــل إلى اتخاذ خرائط الحرارة خطوة أخرى إلى الأمام. فإضافةً إلى إظهار كيفية تداخل البيانات باستخدام اللون؛ تضيف جداول Highlight رقـمًا في الأعلى لتقديــم تفاصيل إضافية. أي إنها جداول ثنائيــة الأبعاد ذات خلايا مملوءة بقيم رقمية وتدرجات لونية. فعلى سبيل المثال: من الممكن عرض أداء مندوبي المبيعات على حسب نوع المنتج وحجم المبيعات.

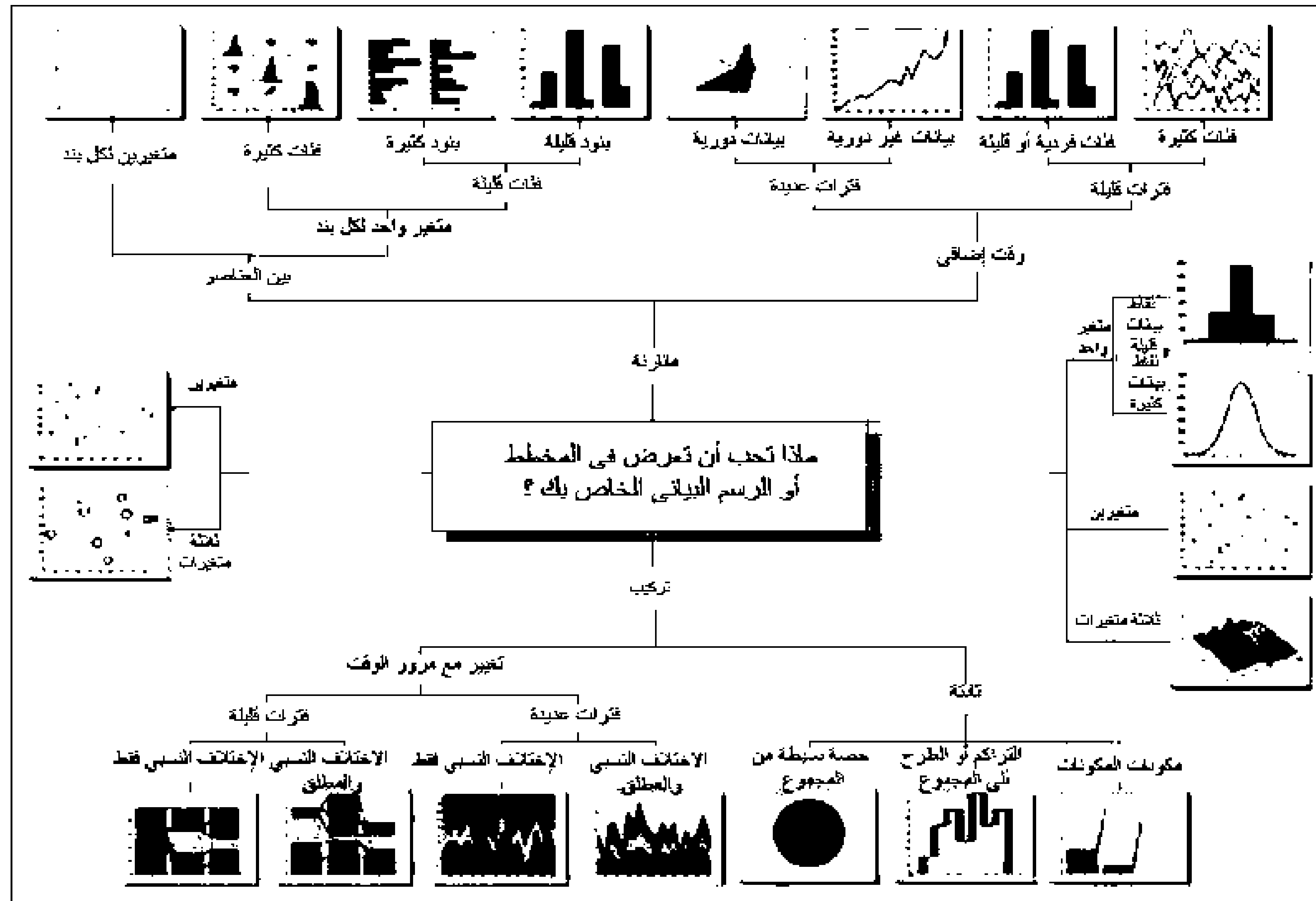
الخرائــط الشــجرية (Tree Map): وهي تعرض البيانات الهرمية (المهيكلة بشــكلٍ شــجري) كمجموعة من المســتطيلات المتداخلة. إذ يتمُّ إعطاء كلِّ فرع من الشــجرة مســتطيلًا، والذي يتم بعد ذلك تقاربه مع مســتطيلات أصغر تمثل الفروع. ويحتوي مستطيل عقدة الورقة على مساحة تتناسب مع بُعدٍ مُحدَّد في البيانات. وغالبًا ما يتم تلوين عُقد الأوراق؛ لإظهار بُعدٍ منفصلٍ للبيانات. وعندما ترتبط أبعاد اللون والحجم بشكلٍ أو بآخر مع هيكل الشجرة؛ فيمكن بسهولة رؤية الأنماط التي يصعب اكتشافها بطرقٍ أخرى، كأن يحظى لون معين بأهمية خاصة. أما الميزة الثانية لخرائط الأشــجار؛ فهي أنها تســتخدم الفضاء بكفاءة؛ وذلك عن طريق الشــكل الهنــدسي الذي تتمتع به. ونتيجةً لذلك؛ فيمكنها عرض آلاف العناصر بشكلٍ مقروء وواضح على الشاشة في وقتٍ واحد.

أيُّ مخططٍ أو رسم بياني يجب أن تستخدمه؟

ما هو الأفضل، فيما شرحناه ســابقاً، المخطط أم الرســم البياني؟ الإجابة سهلة إلى حَدٍّ ما: فلا يُوجَد مخططٌ أو رسمٌ بياني واحد هو الأفضل على الإطلاق؛ فلو كان الأمرُ كذلك لما كان لدينا هذا الكمُّ الكثيرُ من أنواع الرســومات والمخططات البيانيــة. فلدى كلِّ واحدٍ منها «مهارات» مختلفة بعــض الــشيء في تمثيل البيانات عن الآخر. لذلــك؛ ينبغي أن يكون الســؤال الصحيح هو: «أيُّ مخططٍ أو رسم بياني هو الأفضل لمهمة معينة؟» ويمكن أن تساعد إمكانات المخططات الموضَّحة في القســم السابق في اختيار واستخدام الرســم / المخطط البياني المناسب لكلِّ مهمة؛ ولكن يظل من الصعب فرزها. كما يمكن استخدامُ عدة أنواع مختلفة من الرسوم/ المخططات البيانية لنفس مهمة التمثيل التصويري. وتُعَدُّ إحدى القواعد الأساســية هي اختيار واســتخدام أبسط البدائل لتسهيل فهم الجمهور المقصود واستيعابه.

على الرغم من عدم وجود خوارزمية مقبولة وشاملة على نطاقٍ واسعٍ لاختيار وتصنيف المخطط/ الرســم البياني، يُقدِّم الشــكل ٢-٢١ تنظيمًا شــاملًا ومنطقيًّا إلى حدٍّ كبيرٍ لأنواع المخططات/ الرسوم البيانية في بنية شــبه تصنيفية (نُشرت النســخة الأصلية منه في أبيلا ٢٠٠٨). ويجب أن يتمَّ تنظيمُ

البنية التصنيفية بالشـكل الذي يجيب عن السـؤال التالي: «ما الذي ترغبُ في إظهاره في الرسـم/ المخطط البياني؟» بمعنى؛ ماذا سيكون الغرض من الرسم/ المخطط البياني. وعلى هذا المستوى؛ يقوم التصنيف بتقسـيم الغرض إلى أربعة أنواع مختلفة - العلاقة والمقارنة والتوزيع والبنية - كما يُقسِّم الفروع إلى فئات فرعية بناءً على عدد المتغيرات المتضمنة والاعتماد على الوقت في التصوير المرئي.

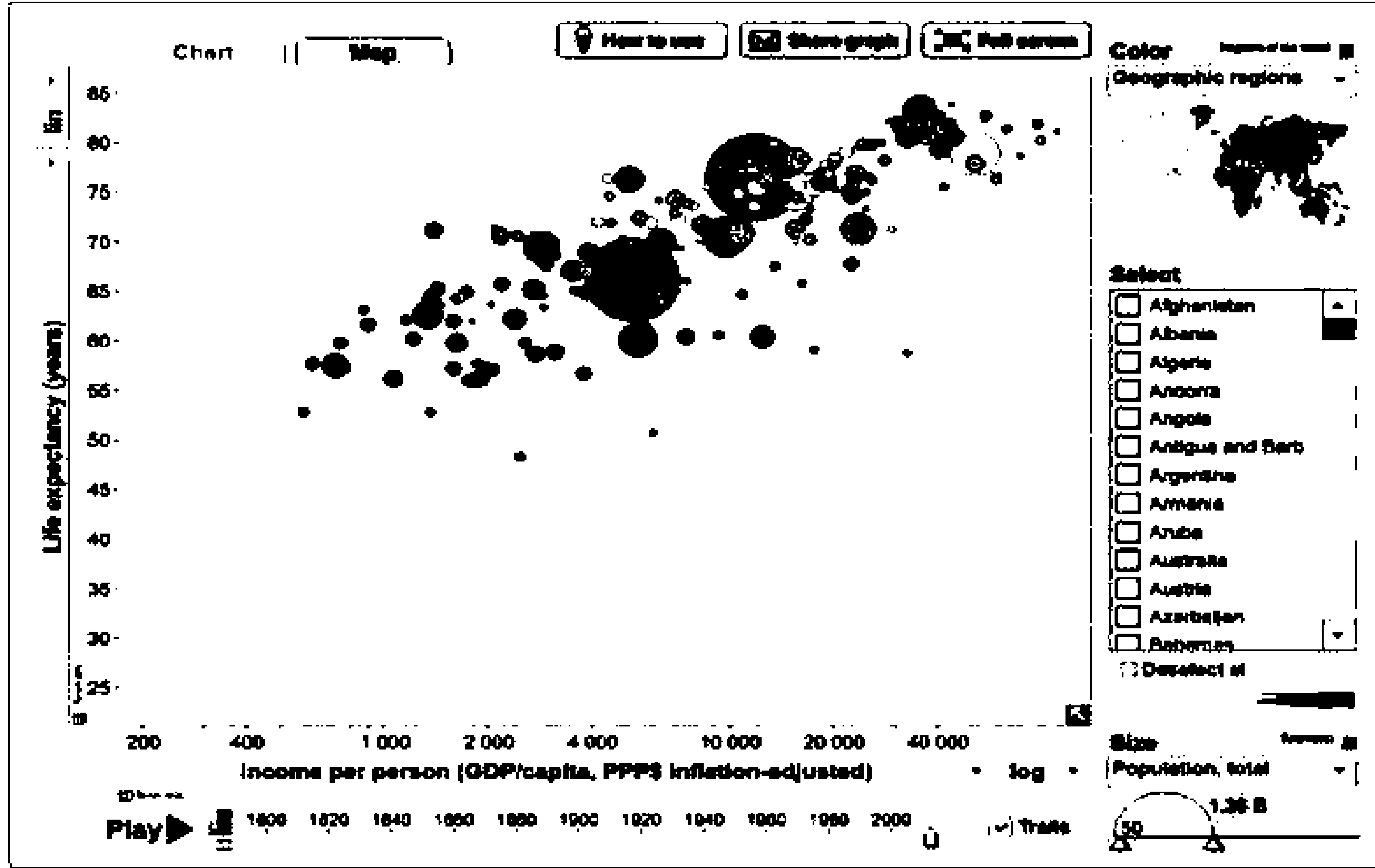


source: Adapted from Abeia. A. (2008) Advanced Presentaion by desighn: Creating Communication that drives action. Newyork: wiley

شكل ٢-٢١: تصنيف الرسوم والمخططات البيانية

على الرغم من أن هذه الرسوم والمخططات البيانية تغطي جزءًا كبيرًا ممَّا هو شائع الاستخدام في تصوير المعلومات؛ فإنها لا تغطي كلَّ ما يتعلق بهذا الموضوع. وفي الوقت الحاضر يمكننا إيجادُ العديـدِ من الرسـوم والمخططات البيانية المتخصِّصة الأخرى التي تخـدم غرضاً معيناً. علاوةً على ذلـك؛ يتمثـل الاتجاه الحالي في دمج/ تهجين وتحريك هذه الرسـوم البيانيـة؛ للحصول على رؤيةٍ أفضـل وأكـثر بديهية لمصادر البيانات المعقدة والمتقلِّبة اليوم. فعلى سـبيل المثال: توفر الرسـوم البيانية من نوع Bubble، والرسوم التفاعلية، والرسوم المتحركة والمتاحة على موقع (Gapminder gapminder.org) وسـيلةً مثيرةً للاهتمام لاستكشاف بيانات الصحة والثروة والسكان من منظورٍ مُتعدِّد الأبعاد. ويوضِّح الشكل ٢-٢٢ أنواع الشاشات المتوفرة في الموقع. ويوضِّح هذا الرسم البياني

كلًا من حجم السكان، ومتوسط العمر المتوقع، ونصيب الفرد من الدخل على مستوى القارة؛ كما يتمُّ عرض رسوم متحركة متغيرة بمرور الوقت توضِّح كيفية تغيُّر هذه المتغيرات بمرور الوقت.



شكل ٢-٢٢: مخطط رسم جغرافي يُظهر ثروات وصحة الأمم

أسئلة مراجعة على القسم ٢-٩:

١- من وجهة نظرك، لماذا يُوجَد هناك العديدُ من الأنواع المختلفة من الرسوم والمخططات البيانية؟

٢- ما هي الاختلافات الرئيسية بين الرسوم الخطية والشريطية والدائرية؟ ومتى يجب عليك استخدامُ أحدهم دون الآخر؟

٣- لماذا تستخدم الخريطة الجغرافية؟ وما هي أنواع الرسوم البيانية الأخرى التي يمكن دمجها مع خريطة جغرافية؟

٤- أوجِد واشرح دور نوعين من الرسوم البيانية التي لم يتم تغطيتها في هذا القسم.

٢-١٠ نشأة التحليلات المرئية:

كما لاحظ سيث غرايمز (a,b 2009)، هناك «ذوق متزايد» من تقنيات وأدوات تصوير البيانات التي تمكِّن مُستخدِمي تحليلات الأعمال وأنظمة ذكاء الأعمال من تحسين «علاقات الاتصال، وإضافة سياق تاريخي، وكَشْف الارتباطات الخفية، وسرد قصص مقنعة توضِّح العمل وتدعو لاتخاذ إجراء».

كـما تؤكد أحدث إصـدارات Magic Quadrant على منصات ذكاء الأعـمال ومِنصات التحليلات الصـادرة عـن Gartner في فبراير ٢٠١٦م على أهمية تمثيل البيانـات في ذكاء الأعمال والتحليلات. وكما يوضِّح الرسـم البيـاني؛ فإن جميع مقترحي الحلول وأصحاب الـرؤى في الشركات القيادية هم إمـا شركات تصوير معلومات حديثة التأسـيس نسـبيًّا (مثل: برمجيـات Tableau أو QlikTech) أو شركات تحليلات كبيرة وراسـخة (مثل مايكروسـوفت (Microsoft)، وSAS، وIBM، وSAP، و Alteryx، وMicroStrategy) والتي تركِّز جهودها بشكلٍ متزايدٍ على تصوير المعلومات والتحليلات المرئية. وتعرض الرؤية الفنية ٢-٢ تفصيلاً لأحدث ما صدر من Magic Quadrant لـ Gartner.

في ذكاء الأعـمال والتحليـلات، تمحورت التحَدِّيات الأساسـية لتصوير البيانـات حول التمثيل البديهي لمجموعات البيانات الكبيرة والمعقدة ذات الأبعاد والمقاييس المتعددة. وبالنسـبة للجزء الأكبر، عادةً ما تتضمَّن المخططات والرسوم البيانية النموذجية والعناصر المرئية الأخرى المُستَخدَمة في هـذه التطبيقات بُعدَين، وأحيانًا ثلاثة أبعاد، ومجموعـات فرعية صغيرة جدًا من مجموعات البيانات. في المقابل؛ فإن البيانات الموجودة في هذه الأنظمة موجودة في مستودع البيانات. وكحد أدنى؛ تتضمَّن هذه المستودعات نطاقاً من الأبعاد (مثل: المنتج، الموقع، الهيكل التنظيمي، الوقت)، ومجموعـة من المقاييس، وملايين من خلايا البيانات. وفي محاولةٍ للتصَدِّي لهذه التحَدِّيات، طوَّر عددٌ من الباحثين مجموعةً متنوعةً من تقنيات التصوير الجديدة.

رؤية فنية ٢-٢

جارتنر ماجيك كوادرانت لمِنصَّات ذكاء الأعمال والتحليلات

يُعَـدُّ جارتنر-إنك المؤسـسَ لشركة ماجيـك كوادرانت، وهـي شركةٌ رائـدةٌ في مجال أبحاث واستشارات تقنية المعلومات للتداول العَلَني في الولايات المتحدة بعائدات سنوية تجاوزت ٢ مليار دولار في عـام ٢٠١٥م. وقد أُنشِـئَت في عـام ١٩٧٩م، وكان لديها آنذاك ٧٦٠٠ شريك، بما في ذلك ١٦٠٠ محلل واستشاري أبحاث والعديد من العملاء في ٩٠ دولة. ويُعَدُّ ماجيك كوادرانت أسلوبَ بحـثٍ قام بتصميمه وتنفيذه جارتـنر لمتابعة وتقييم التقدُّم الحاصل ومراكز الشركات في سـوق مُحـدَّدة قائمة على التقنية. ومن خلال تطبيق المعالجة الرسـومية ومجموعة مُوحَّدة من معايير التقييم، وتساعد ماجيك كوادرانت المستخدمين على فهم كيفية تقييم مُقدِّمي التقنية في السوق.

قامـت غارتنر بتغيير اسـم ماجيك كوادرانـت من «منصات ذكاء الأعـمال» إلى «منصات ذكاء الأعمال والتحليلات»؛ للتأكيد على الأهمية المتزايدة لقدرات التحليلات لنظم المعلومات التي تبنيها

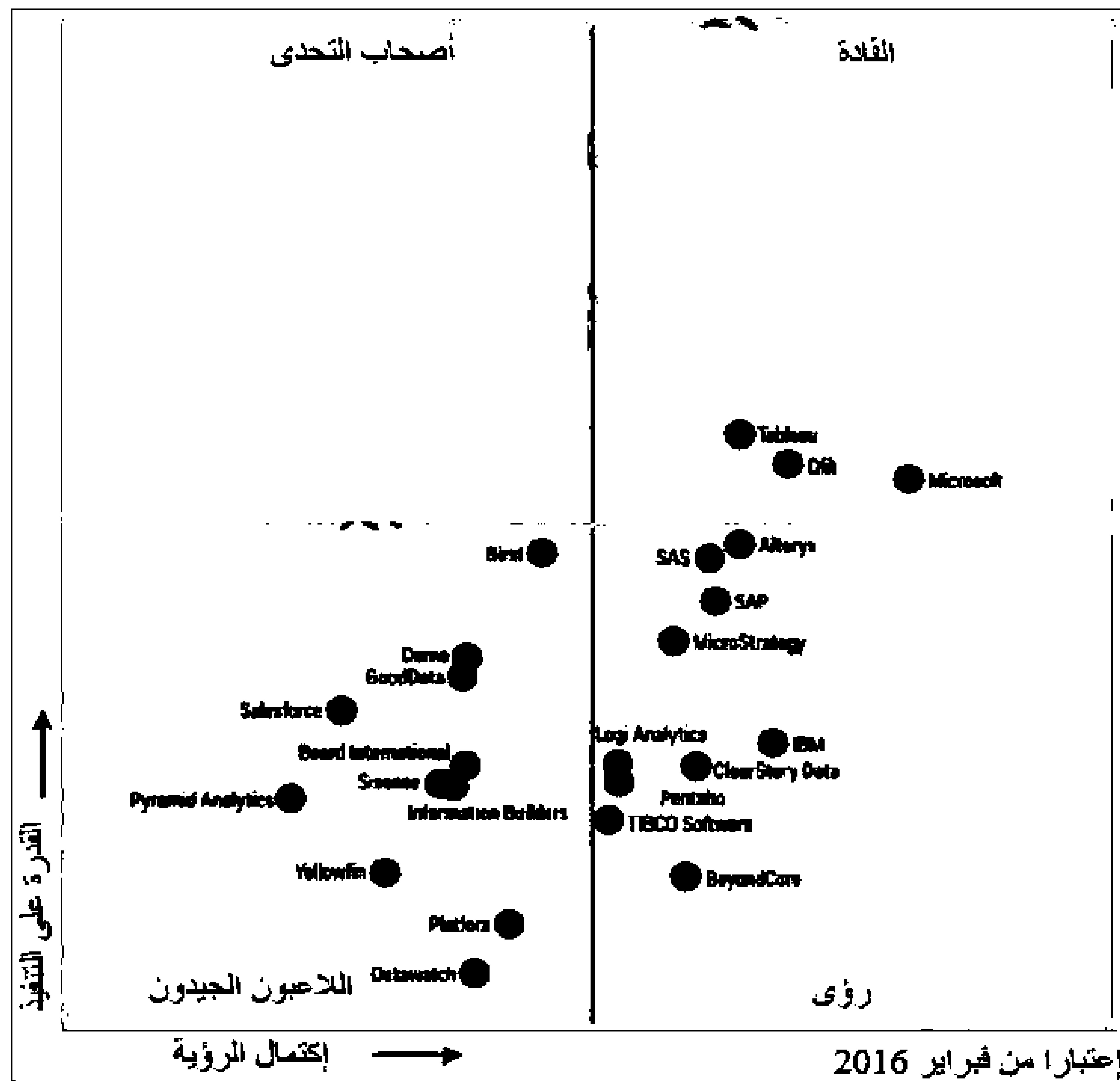
المنظـمات الآن. تحَدِّد جارتنر سـوقاً لمنصة ذكاء الأعمال والتحليـلات كمنصة برمجية تقوم بتوفير ١٥ إمكانية عبر ثلاث فئات، وهي: التكامل، وتسـليم المعلومات، والتحليل. تُمكّن هذه الإمكانيات المؤسسات من بناء أنظمة دقيقة للتصنيف والقياس؛ بغرض دعم صُنع القرار وتحسين الأداء.

ويُوضِّح الشكل ٢-٢٣ أحدث نظام لماجيك كوادرانت الخاص بمنصات ذكاء الأعمال والتحليلات. إذ تضـع ماجيك كوادرانت مُقدِّمي الخدمات في أربع مجموعات (اللاعبين المتخصِّصين، وأصحاب التحَـدِّي، والحالمين، والقـادة) على محورين هما: اكتمال الرؤية (المحور السـيني) والقدرة على التنفيـذ (المحور الصـادي). وكما يظهر كوادرانت بشـكلٍ واضح؛ فإن معظـم مُقدِّمي BI/BA المعروفـين جيدًا يتـم وضعهم في فئة «القادة»، في حين يتمُّ وَضْـع العديد من مُقدِّمي الخدمات الناشئة الأقل شهرةً والجديدة نسبيًّا في فئة «اللاعبين المتخصصين».

يبدو أنَّ التحول في سـوق منصـة ذكاء الأعمال والتحليلات مُتعدِّد السـنوات من التقارير التي تقودها تقنية المعلومات إلى تحليلات الخدمات الذاتية التي تقودها الشركات قد اجتاز نقطة التحول. إن معظم عمليات الشراء الجديدة هي من منصات التحليلات المرئية الحديثة، المرتكزة على الأعمال التجاريـة؛ مما يؤدي إلى فرض منظور سـوقي جديد، يعمل على إعادة ترتيـب وجهة نظر البائعين بشـكلٍ ملحوظ. إنَّ معظم النشـاط في سـوق منصة ذكاء الأعمال والتحليلات هو من المؤسسـات التي تحاول رفع قدراتها التصويرية والانتقال من النسـق الوصفي إلى مستويات التحليلات التنبؤية والمفروضة. وقد ركَّز البائعون في السـوق بشـكلٍ كبيرٍ على تلبية طلب المسـتخدِم. وإذا كان هناك سـوق واحد في عام ٢٠١٥م؛ فسيصبح اكتشـاف/ تصوير البيانات هو فن الهندسة المعمارية السائد. وفي الوقت الذي يعمل فيه مُورِّدو اكتشاف/ تصوير البيانات، مثل: Tableau وQlik ومايكروسوفت (Microsoft) على ترسـيخ مراكزهم كقادة لكوادرانت، يحاول آخرون (سـواءً من مُقدِّمي الحلول/ الأدوات الناشئين والكبار أو الراسخين) الانتقال من طور أصحاب الرؤى إلى طور القادة.

يستمر هذا التركيز على اكتشاف/ تصوير البيانات من معظم القادة وأصحاب الرؤى في السوق - التي تعمل الآن على تعزيز الأدوات مع تكامل البيانات سهلة الاستخدام من جانب الشركات، إلى جانب التخزين المُدمَج وطبقات الحوسـبة والحفر غير المقيَّد - في تعجيل الاتجاه نحو اللامركزية وتمكين المسـتخدِم من ذكاء الأعمال والتحليـلات، ويُعضِّد إلى حَدٍّ كبيرٍ قدرة المنظمات على إجراء تحليلات تشخيصية.

Source: Gartner Magic Quadrant, released on February 4, 2016, gartner.com (accessed August 2016).



شكل ٢-٢٣: Magic Quadrant لمنصات ذكاء الأعمال والتحليلات

التحليلات المرئية:

تُعرَّف التحليلات المرئية بأنها عبارة عن مصطلحٍ تمَّ صياغته مؤخرًا وغالبًا ما يُستَخدَم بشكلٍ لا يعني شيئًا أكثر من التصوُّر المعلوماتي. والمقصود بـالتحليلات المرئية؛ هو مزيجٌ من تصوير البيانات والتحليــلات التنبؤيــة. ففي حين يهدف تصويرُ المعلومات إلى الإجابة عن الأســئلة التالية: «ماذا حــدث؟» و«ما الذي يحدث؟» ويرتبط ارتباطًا وثيقًا بــذكاء الأعمال (التقارير الروتينية، وبطاقات الأداء، ولوحات المعلومات)، في حين تســتهدف التحليلات المرئية الإجابة عن الأسئة: «لماذا يحدث هذا؟» و«ما هو الاحتمال الأكبر أن يحدث؟» وعادةً ما يرتبط بتحليلات الأعمال (التوقع، التقسيم إلى قطاعــات، تحليــل الارتباط). ويضيف العديدُ مــن مُورِّدي تصوير المعلومــات القدرات التي تجعلنــا نطلق عليهم اســم مُوفِّري حلول التحليلات المرئية. يقــترب معهد SAS - وهو أحد كبار مــزوِّدي حلول التحليــلات منذ فترة طويلة - من اتجاهٍ آخر؛ فهم يدمجــون قدراتهم التحليلية في

بيئة تصويـر بيانات عالية الأداء والتي يُطلقون عليها التحليلات المرئية. ولا يختلف إعدادُ تقارير الأعـمال كثيرًا عن سَرْد القصة من حيث كونه مرئيًا أو غير مرئي، آليًا أو يدويًا، عبر الإنترنت أو على الورق. وتعرض الرؤية الفنية ٢-٣ وجهة نظر مختلفة وغير تقليدية لتحسين تقارير الأعمال.

رؤية فنية ٢-٣

سرد قصص عظيمة باستخدام البيانات والتصوير المرئي

يمتلـك كلُّ مَنْ لديه بيانات لتحليلها قصصًا ليرويها، سـواء كانت تُشـخِّص أسـباب الخلل في التصنيـع، أو تبيع فكـرةً جديدةً بطريقة تعكس خيـال الجمهور المُسـتَهدَف، أو تعطي الزملاء فكرةً عن برنامج تحسـين خدمة عملاء معين. وعندما يتمَخَّض سَرْد قصة ما عن خيارٍ إستراتيجي كبـير يمكِّنـك أنت وفريق الإدارة العليـا التابع لك من اتخاذ قرار قوي؛ فإنَّ تقديم قصة تسـتندُ إلى الواقـع قـد يكون أمرًا صعبًا للغاية. وفي جميع الحـالات؛ فهي مهمةٌ كبيرة. فعليك أن تكون القصـة مثـيرةً للاهتمام وعالقةً بالذاكرة. فأنت تعلم أنك بحاجة إلى أن تُبقي الأمر بسـيطًا على مديريك التنفيذيين والزملاء المشغولين. ولكنك تعلم أيضًا أنه ينبغي لك أن تكون واقعيًا وموجِّهًا بالتفاصيل، وأن تكون مشغولًا بالبيانات، لا سيما في عالم اليوم الذي يركِّز على المقاييس.

إن لتقديـم البيانـات والحقائق متعةً فائقـةً، ولكن عندما يتمُّ إغراق الزمـلاء والإدارة العليا بالبيانات والوقائع دون سـياق؛ فذلك يُعَدُّ خسـارةً بلا شـك. لقد قمنا جميعًـا بتجربة عروضٍ تقديميةٍ باسـتخدام طبقات الشرائح الكبيرة، فقط لنكتشف أن الجمهور غارق في البيانات لدرجة أنهم لا يعرفون ما يفكِّرون به، أو أنهم بالكامل لا يأخذون سوى جزءٍ بسيطٍ من النقاط الأساسية.

ابدأ في إشراك فريقك التنفيذي، وشَرْح إستراتيجياتك ونتائجك بقوةٍ أكبر من خلال الاقتراب من مهمتك كراوي. ستحتاج إلى إجابة سؤال: «ما هي قصتك؟» (الحقائق والبيانات)؛ ولكنك ستحتاجُ أيضًا إلى أسـئلةٍ من نوع: «من؟» و«كيف؟» و«لماذا؟» والسـؤال الذي يفتقده الجمهور غالبًا هو: «ثم ماذا؟»، إنها عناصر القصة التي سـتجعل بياناتك ذات صِلة وملموسـة لجمهورك. إن إنشاء قصة جيدة من شأنه مساعدة الإدارة العليا في التركيز على ما هو مهم.

لماذا قصة؟

القصـص تنفـخُ الروح في الحقائق والبيانـات وتجعلها حيةً. كما يمكن أن تسـاعدك القصة على فهم البيانات وترتيبها من خلال مجموعةٍ متباينةٍ من الحقائق. فهي تجعل من السـهل تذكُّر النقاط

الأساسية، ويمكن أن ترسمَ صورةً حيةً لما يمكن أن يبدو عليه المستقبل. كما تعمل القصصُ أيضًا على خَلْق نوعٍ من التفاعل؛ بحيث يضع كلُّ واحدٍ من الجمهور نفسه في القصة، ويبدأ التعامل مع الموقف.

تَســتَخدم الثقافاتُ منذ زمن طويل الــسرد القصصي؛ لتمرير المعرفة والمحتــوى. ويُعَدُّ سردُ القصــص في بعــض الثقافات أمرًا حيويًّا لهويتهم. فعلى ســبيل المثــال: في نيوزيلندا؛ يقوم بعض شعوب الماوري بوشم وجوههم بالموكوس. والموكوس هو وشمٌ للوجه يحتوي على قصة عن قدماء القبائــل والعائلات. قد يكون لدى الرجل تصميمٌ وشــمي على وجهــه يُظهِر ملامح رأس مطرقة لتســليط الضوء على الصفات الفريدة المتعلقة بنَسَبه. ويُعبِّر التصميمُ الذي يختاره عن جزء من «ذاته الحقيقية» ووطن أسلافه.

وبالمثل، عندما نحاول أن نفهمَ قصةً ما، يتنقل راوي القصة لإيجاد الشــمال الحقيقي «True North». وإذا كانت الإدارة العليا تتطلَّع إلى مناقشــة كيفية اســتجابتها في المســتقبل للتغييرات التنافســية؛ فإن القصة الجيدة يمكِن أن تكون ذات معنًى ومُنقَّاة من الكثير من الضوضاء. فعلى ســبيل المثال: قد يكون لديك حقائق وبيانات من دراســتين؛ إحداهما تتضمَّن نتائج من دراســة إعلانيــة والأخــرى من دراســةٍ حول الرضا عن المنتــج. إن تطوير قصة لما قمت بقياســه في كلتا الدراستين يمكن أن يساعد الأشخاص على الحصول على رؤية كلية في الوقت الذي تُوجَد فيه أجزاء متباينة. ولجَمْع المُوزِّعين حول منتجٍ جديد، يمكنك اســتخدامُ قصة لإعطاء رؤيةٍ لما يمكن أن يبدو عليه المســتقبل. والأهمُّ مــن ذلك أن السَّرد القصصي تفاعليٌّ فعادةً ما يَســتخدم مُقدِّم البرامج كلــماتٍ وصوراً يمكن لأفــراد الجمهور وَضْع أنفســهم فيها. ونتيجةً لذلك؛ يصبحــون أكثر تعلُّقًا بالمعلومات وأفضل فهمًا لها.

إذًا ما هي القصة الجيدة؟

يُمكِــن لمعظم الناس بســهولةٍ حِفظُ وترديدُ نصِّ فيلمهم/ كتابهــم المفضل. أو يتذكرون قصةً مضحكةً شــاركها أحد الزملاء مؤخرًا. لماذا يتذكر الناسُ هذه القصص؟ لأنها تحتوي على صفاتٍ مميــزةً معينة. أولًا: تحتوي القصة الجيدة على شــخصيات عظيمة. وقــد يكون لدى القارئ أو المُشــاهِد في بعــض الحالات خــبرةٌ غير مباشرة عندما يصبح مشــاركًا في الشــخصية. والتي (أيْ: الشخصية) تواجه بعد ذلك تحديًا صعبًا؛ ولكنه قابلٌ للتصديق. ويجب أن تعترض طريق الشخصية بعــض عقبــات؛ ولكنها تتخطاها. وأخــيرًا؛ وبنهاية القصة تكون النتيجة واضحــةً أو التخمين بها ممكن. وقد لا يتم حل الموقف - ولكن لا بُدَّ للقصة من نقطة نهاية واضحة.

فكِّر في تحليلك كقصة - استخدم هيكل قصة:

عند صياغة قصة غنية بالبيانات؛ يكون الهدفُ الأول هو العثور على القصة. مَنْ هم الشخصيات؟ ما هي الدراما أو التحَدِّي؟ ما هي العقبات الواجب تخطِّيها؟ وفي نهاية قصتك؛ ما الذي تريد من جمهورك أن يفعله كنتيجة لذلك؟ وبمجرد معرفة لُبِّ القصة، قُمْ بتركيب باقي عناصر قصتك؛ أيْ: حَدِّد شخصياتك، افهم التحَدِّي، وحَدِّد العقبات، ثم قم ببلورة النتيجة أو سؤال القرار. تأكَّد من أنك واضحٌ بما تريد أن يفعله الناس كنتيجة لذلك. وسيُجسِّد هذا كيفية استرجاع الجمهور لقصتك. ثم باستخدام عناصر القصة، قم بكتابة القصة المصوَّرة، والتي تمثِّل هيكل وتكوين قصتك. وعلى الرغم من أن تخَطِّي هذه الخطوة ممتعٌ للغاية؛ فإنه من الأفضل أولًا فهم القصة التي ترويها، ومن ثم التركيزُ على بنية وشكل العرض. وبمجرد وَضْع لوحة العمل في مكانها، ستقع بقية العناصر الأخرى في مكانها كذلك. وستساعدك لوحة العمل على التفكير في أفضل المجازات أو الاستعارات، وعلى تحديد التحَدِّي أو الفرصة بوضوح. وأخيرًا، ستساعدك على رؤية التدفق والتحوُّلات اللازمة. وستساعدك لوحة العمل أيضًا على التركيز على العناصر المرئية الأساسية (الرسوم، والمخططات البيانية، والجرافكس) التي تحتاج إلى استرجاع بواسطة الموظفين الإداريين لديك. ويُوضِّح الشكل ٢-٢٤ قصة تأثير القروض الصغيرة من وجهة نظر عالمية داخل بيئة التحليلات المرئية في Tableau.

باختصار، لا تخف من استخدام البيانات لسَرْد قصص عظيمة. ومن الأهمية بمكان أن تكون القصة واقعيةً وقائمةً على التفاصيل وموجهةً بالبيانات في العالم المرتكز على القياس في هذه الأيام؛ ولكن ليس بالضرورة أن تكون مملًا وطويلًا. وفي الواقع، يمكنك من خلال إيجاد القصص الحقيقية في بياناتك ومتابعة أفضل الممارسات، كَسْب أشخاص للتركيز على رسالتك - وبالتالي على ما هو مهم. وفيما يلي أفضل تلك الممارسات:

١- فكِّر في تحليلك كقصة، استخدم هيكل قصة.

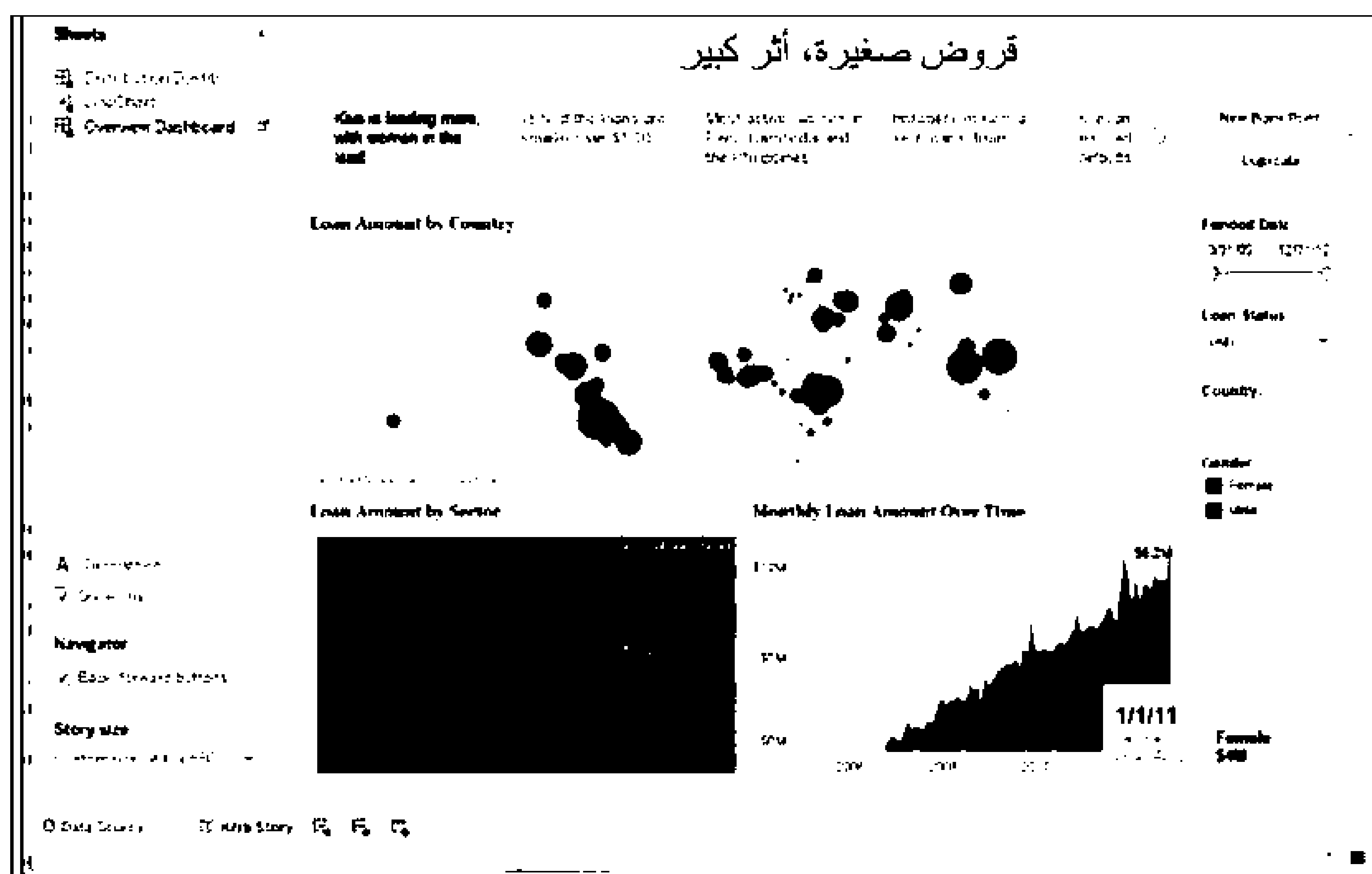
٢- كُنْ حقيقيًّا، تتدفق قصتك.

٣- كُنْ مرئيًّا، فكِّر في نفسك كمؤلف أفلام.

٤- اجعل الأمرَ سهلًا على جمهورك وعلى نفسك.

٥- وَجِّه دعوة، واجرِ مناقشةً مباشرة.

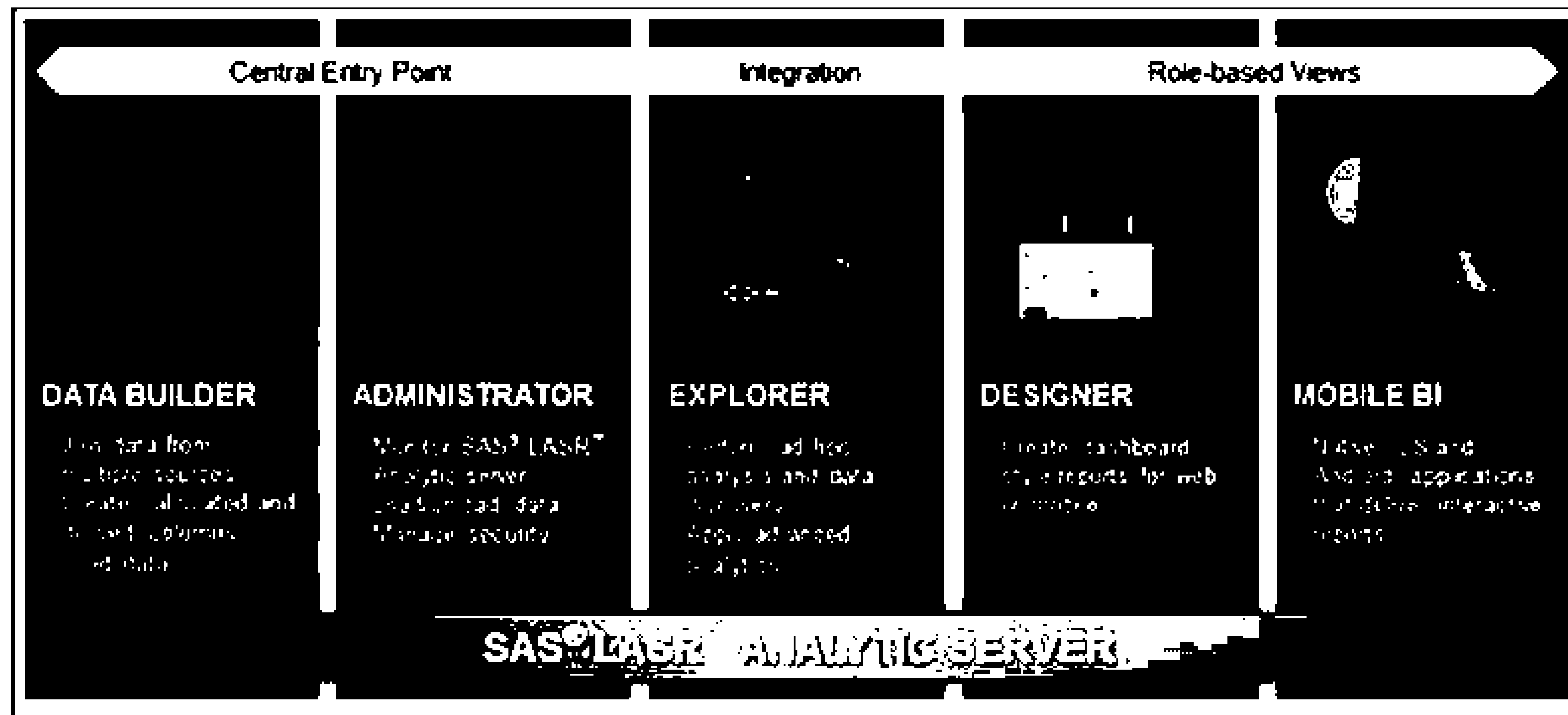
Source: Fink, E., & Moore, S. J. (2012). Five best practices for telling great stories with data. White paper by Tableau Software, Inc., www.tableau.com/whitepapers/telling-data-stories (accessed May 2016).



شكل ٢-٢٤: تصوير مرئي لقصة في برنامج Tableau

بيئات تحليلات بصرية عالية القوة:

نظـرًا للطلبِ المتزايدِ على التحليلات المرئية المقترنة بأحجـام البيانات السريعة النمو؛ فهناك تحرُّك حثيث باتجاه الاستثمار في أنظمة تصوير البيانات ذات الكفـاءة العالية. ومع تحرُّكاتهم الأخـيرة في مجال التحليلات المرئية؛ أصبح معهد SAS العملاق للبرمجيات الإحصائية الآن من بين أولئك الذين يقودون هذه الموجة؛ إذ يُعَدُّ منتجهم الجديد، تحليلاتSAS المرئية، حوسـبةً عالية الأداء، وحَلاً مختزناً في الذاكرة؛ لاستكشاف كميات هائلة من البيانات في وقتٍ قصيرٍ للغاية (فوري تقريبًا). فهي تُمكِّن المستخدمين من تحديد الأنماط، وتسليط الضوء على الفرص المؤهلة لمزيدٍ من التحليـل، ونقل نتائج مرئية عبر تقارير الويب أو مِنصة متحركة، مثل: الأجهزة اللوحية والهواتف الذكية. ويُوضِّح الشـكل ٢-٢٥ البنية العالية المسـتوى لِمنصة تحليلات SAS المرئية. ويُوجَد على أحد طرفي الهندسة المعمارية، قدرةٌ عالميةٌ على إنشاء البيانات وإمكانيات المسؤول؛ مما يؤدي إلى مُستكشِف، ومُصمِّم تقارير، ووحدات ذكاء الأعمال محمولة قيِّمة جدًا؛ مما يوفر بشكلٍ جماعي حلاً تحليلياً مرئياً من البداية إلى النهاية.



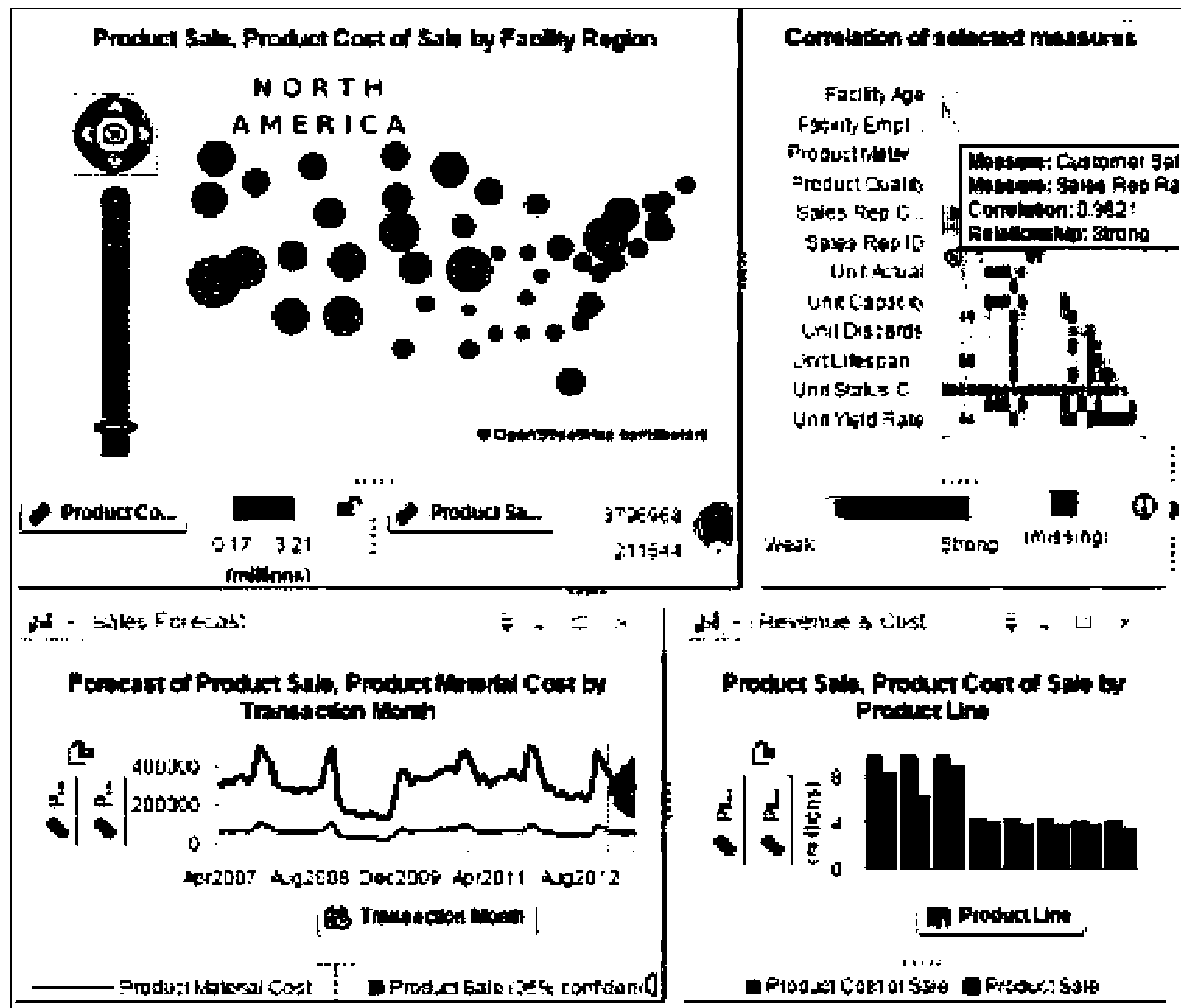
شكل ٢-٢٥: نظرة عامة على بنية تحليلات SAS المرئية

وفيما يلي بعضُ المزايا الرئيسة المقترحة بواسطة تحليلات SAS المرئية:

- تُمكِّن جميع المستخدمين من استخدام تقنيات استكشاف البيانات والتحليلات سهلة الإجراء؛ من أجل تحسين عملية صُنْع القرار. كما تُمكِّن تحليلات SAS المرئية أنواعًا مختلفةً من المستخدِمين من إجراء عمليات استكشــاف سريعة وشاملة لجميع البيانات المتاحة. ولا يُفضَّل أخذ عينات لتقليل حجم البيانات.
- تعمل واجهات الويب التفاعلية ســهلة الاســتخدام على توسيع جمهور التحليلات، وتمكينه من تحصيل رؤى جديدة. كما يُمكِن للمستخدمين الاطلاع على مزيدٍ من الخيارات، واتخاذ قرارات أكثر دقة، والوصول للنجاح بشكلٍ أسرع من ذي قبل.
- تُســهِّل الإجابة عن الأســئلة المعقدة بشكلٍ أسرع، وتعزيز مســاهمات موهبتك التحليلية؛ إذ تدعم تقنية تحليلات SAS المرئية عملية اكتشــاف البيانات واستكشافها من خلال توفير نتائج فائقة السرعة؛ لتمكين إجراء تحليل أفضل وأكثر تركيزًا. ويمكِن للمســتخدمين المحنَّكين تحليليًّا تحديــدُ مجالات الفرص أو التهديــدات المتعلقة بالكميات الكبيرة مــن البيانات؛ بحيث يمكن تحقيق مزيدٍ من التحليلات بسرعة.
- تُحسِّــن مشــاركة المعلومات والتعاون. فيُمكِن لأعداد كبيرة من المســتخدمين، بما في ذلك ذوو المهارات التحليلية المحدودة؛ عَرْضُ التقارير والرســوم البيانية والتفاعل معها بسرعة عبر الويب وملفــات Adobe PDF وأجهزة iPad المحمولة؛ في حين تحافظ تقنية المعلومات على التحكُّم في البيانات الأساســية والأمان. وبذلك توفر تحليلات SAS المرئية المعلومات المناسبة للشخص المناسب وفي الوقت المناسب لتحسين الإنتاجية والمعرفة التنظيمية.

- تُحرِّر تقنية المعلومات من خلال مَنْح المستخدمين طريقةً جديدةً؛ للوصول إلى المعلومات التي يحتاجــون إليها. وأيضًا تحرير تقنية المعلومات من وابل المطالب المســتمر من المســتخدمين الذيــن يحتاجون إلى الوصول إلى كمياتٍ مختلفة من البيانات، ووجهات نظر مختلفة للبيانات، وتقارير مُعدَّة لغرضٍ معين، وطلبات المرة الواحدة للحصول على المعلومات.
- تُمكِّــن تقنية المعلومات مــن تنزيل البيانات وإعدادها بســهولةٍ لعِدَّة مســتخدمين. فبمجرد تنزيل البيانات وجَعْلها متاحةً؛ يمكن للمستخدمين استكشاف البيانات بحيوية وإنشاء التقارير ومشاركة المعلومات بمفردهم.
- تُوفِّر مســاحةً للنموِّ بوتيرة مُحدَّدة ذاتيًّا. كما توفر تحليلات SAS المرئية خيار استخدام أجهزة قاعدة البيانات للسلع المُقدَّمة من EMC Greenplum وتيراداتا. وقد تمَّ تصميمه من الألف إلى الياء؛ لتحسين الأداء وقابلية التوسُّع لتلبية احتياجات أيِّ منظمة بأيِّ حجم.

ويعرض الشكل ٢-٢٦ لقطة شاشة لمنصة تحليلات SAS المرئية؛ إذ يتمُّ تصوير التنبؤ بالسلاسل الزمنية وفاصل من الثقة حول التوقعات.



شكل ٢-٢٦: لقطة شاشة من تحليلات SAS المرئية

أسئلة مراجعة على القسم ٢-١٠:

١- ما الأسبابُ الرئيسة وراء النشأة الحديثة للتحليلات المرئية؟

٢- انظر إلى Magic Quadrant لِمنصَّات ذكاء الأعمال والتحليلات. ماذا ترى؟ ناقش وعلل ملاحظاتك.

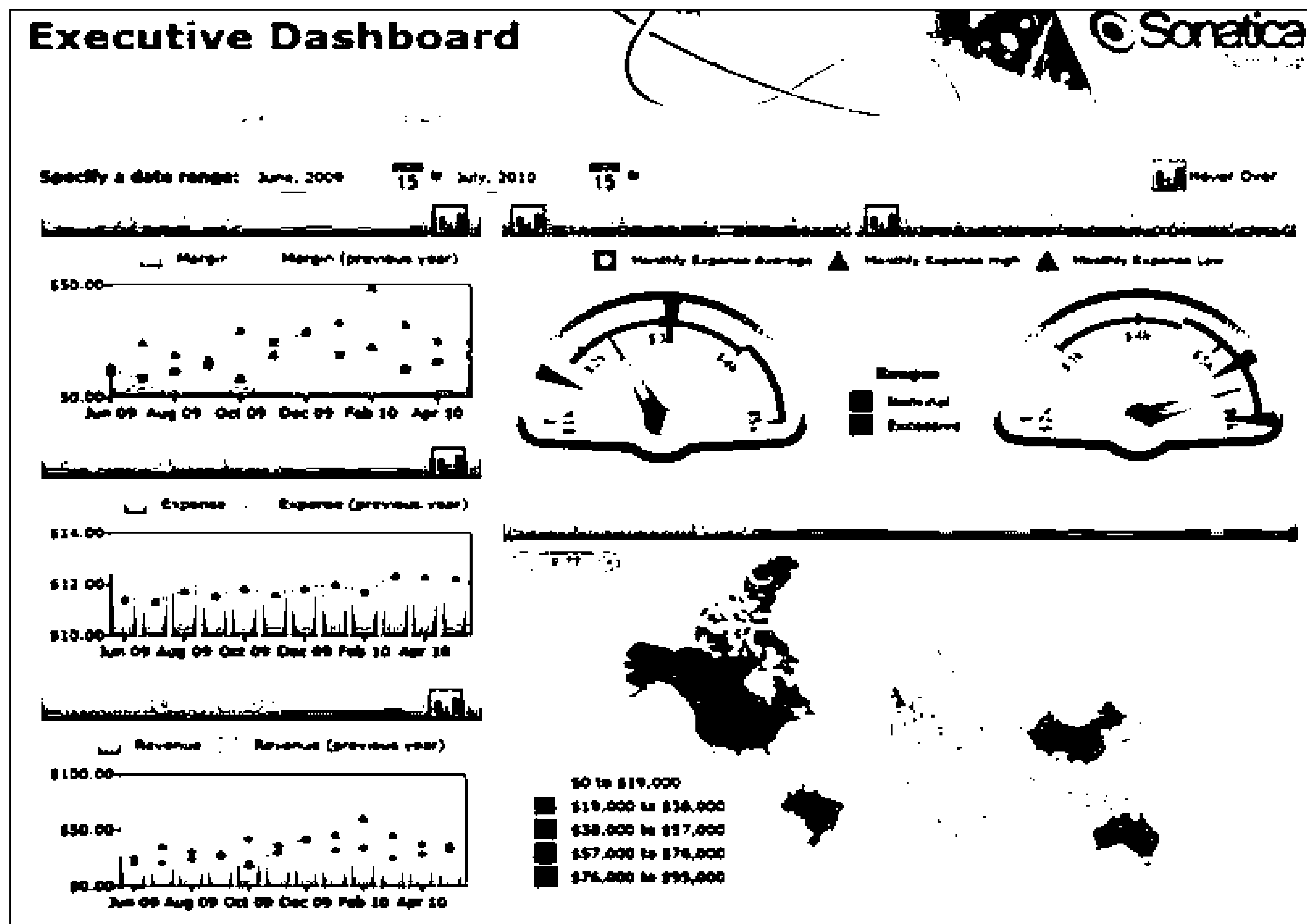
٣- ما الفرق بين تصوير المعلومات والتحليلات المرئية؟

٤- لماذا يجب أن يكون سرد القصص جزءًا من إعدادك للتقارير وتصويرك للبيانات؟

٥- ما هي بيئة التحليلات المرئية عالية الأداء؟ ولماذا نحتاجها؟

٢-١١ لوحات المعلومات:

تُعَدُّ لوحات المعلومات من المكوِّنات الشائعة لمعظم مِنصات ذكاء الأعمال أو تحليلات الأعمال، إن لم يكن لها جميعًا، وكذلك أنظمة إدارة أداء الأعمال، ومجموعة برامج قياس الأداء. توفِّر لوحات المعلومات عروضاً مرئيةً لمعلومات مهمة يتمُّ توحيدها وترتيبها على شاشة واحدة؛ بحيث يُمكِن استيعابُ تلك المعلومات في لمحةٍ واحدةٍ وفحصها بسهولة واستكشافها بشكلٍ أكبر. تظهر لوحة التحكُّم النموذجية في الشكل ٢-٢٧. وتعرض لوحة المعلومات التنفيذية هذه مجموعةً متنوعةً من مؤشرات الأداء الرئيسة (KPIs) لشركة برمجيات افتراضية تُسمَّى Sonatica (لبيع الأدوات الصوتية). كما تُقدِّم تلك اللوحة عرضًا عالي المستوى للمجموعات الوظيفية المختلفة المحيطة بالمنتجات؛ بدءًا من نظرة عامة إلى جهود التسويق والمبيعات والتمويل وإدارات الدعم. ويهدف كلُّ هذا إلى إعطاء صانعي القرار التنفيذيين فكرةً سريعةً ودقيقةً عمَّا يجري داخل المنظمة. على الجانب الأيسر من لوحة المعلومات؛ يمكننا أن نرى (بطريقة السلاسل الزمنية) التغييرات ربع السنوية في الإيرادات والنفقات وهوامش الربح، وكذلك مقارنةُ هذه الأرقام بالأرقام الشهرية للسنوات السابقة. في الجانب العلوي الأيمن؛ نرى وجهين مع مناطق مشفرة بالألوان يعرضان مقدار النفقات الشهرية لخدمات الدعم (الاتصال الهاتفي على اليسار) ومقدار النفقات الأخرى (الاتصال الهاتفي على اليمين). كما يشير الترميز اللوني، وعلى الرَّغم من أن نفقات الدعم الشهرية تقع ضمن النطاقات العادية؛ فإن النفقات الأخرى تقع في المنطقة الحمراء؛ مما يشير إلى القيم المفرطة. وتوضِّح الخريطة الجغرافية في أسفل اليمين توزيع المبيعات على مستوى البلد في جميع أنحاء العالم. تُوجَد خلف هذه الرموز الرسومية مجموعةٌ متنوعةٌ من الوظائف الرياضية التي تجمع العديدَ من نقاط البيانات إلى أعلى مستوياتها من الأرقام ذات المعنى. من خلال النقر على هذه الرموز الرسومية، يُمكِن لمستخدمي هذه المعلومات الانتقال إلى مستويات أكثر دقة من المعلومات والبيانات.



شكل ٢-٢٧: عينة لوحة معلومات تنفيذية

يتمُّ استخدامُ لوحات المعلومات في مجموعةٍ متنوعةٍ من الأنشطة التجارية لمجموعة متنوعة من الأسـباب. فعلى سبيل المثال: في الحالة العملية ٢-٧ ستجد ملخص التنفيذ الناجح لمعلومات لوحات المعلومات من قِبل فريق دالاس كاوبويز لكرة القدم.

حالة عملية ٢-٧

فريق دالاس كاوبويز يحقق نتائج كبيرة مع تابلوه (Tableau) وتكنيون (Teknion)

تأسَّـس دالاس كاوبويز في عام ١٩٦٠م، وهـو فريق كرة قدم أميركي محترف ومقره في إيرفينـغ، تكسـاس. ويتبع الفريق عددٌ كبيرٌ من الأنديـة الوطنية، والتي ربما تكون صاحبة أفضل تمثيل بسجل اتحاد كرة القدم الأميركي لعددِ المباريات المتتالية في الملاعب التي تمَّ بيعها بالكامل.

التحَدِّي:

كان بيـل برياكوس - الرئيس التنفيذي للعمليات في القسـم التجاري بفريق دالاس كاوبويز - وفريقه بحاجةٍ إلى رؤيةٍ أكثر وضوحًا في بياناتهم؛ حتى يتمكنوا من تشـغيلها

بشــكلٍ أكثر ربحية. فتمَّ اختيارُ مايكروســوفت كمنصة أساســية لهذه الترقية بجانب عــددٍ آخر مــن تطبيقات المبيعات والخدمات اللوجســتية والتجــارة الإلكترونية (لكلٍّ ميجــاوات). وتوقَّــع كاوبويز أن بنية المعلومات الجديدة هذه ســوف توفر التحليلات والتقارير اللازمة. ولكن لسوء الحظ لم يحدث هذا، ومن ثم بدأ البحث عن أداة قوية للتحكُّم في المعلومات والتحليلات وإعداد التقارير لسَدِّ هذه الفجوة.

الحل والنتائج:

قام كلٌّ مــن Tableau وTeknion بتوفير إمكانيات إعداد تقارير لوحة المعلومات الفوريــة، والتــي تجاوزت متطلبــات كاوبويز. وبشــكلٍ مُنظَّم وممنهــج، عمل فريق Teknion جنبًا إلى جنب مع مالكي ومســتخدمي البيانات داخل دالاس كاوبويز لتقديم كل الوظائــف المطلوبــة، في الوقت المحَــدَّد وفي حدود الميزانية المرصــودة. وقال بيل لويــزي - نائب الرئيــس في Teknion: «في وقتٍ مبكر من هــذه العملية؛ كنَّا قادرين على الحصول على فهمٍ واضحٍ لما قد يتطلبه الأمر لتشــغيل عملية أكثر ربحيةً بالنســبة لكاوبويز، وتُعَدُّ هذه الخطوة من العملية خطوةً رئيسةً في نهج Teknion مع أيِّ عميل، ودائمًــا ما يقوم بتوزيع أرباح هائلة كلما أحرزت خطة التنفيذ تقدمًا». وأضاف لويزي: «بــكلِّ تأكيد؛ لقد عمل Tableau عن كَثَب معنا ومــع كاوبويز خلال المشروع بأكمله. وقــد حرصنا معًا على التأكد من أن كاوبويز يمكنهم تحقيق أهدافهم في إعداد التقارير والتحليل في زمن قياسي».

والآن؛ ولأول مرة يســتطيع فريقُ دالاس كاوبويز مراقبة أنشــطته التجارية بالكامل بدءًا من التصنيع ووصولًا إلى العميل النهائي، وليس فقط رؤية ما يحدث عبر دورة حياة النشــاط؛ بل التعمُّق أكثر في ســبب حدوثها. واليوم، يتمُّ اســتخدام حلِّ ذكاء الأعمال؛ لإعداد تقرير عن أنشــطة الأعمال للقســم التجاري وتحليلها، والذي يُعَدُّ المسؤول عن جميــع مبيعات العلامة التجارية في دالاس كاوبويز. وتشــير تقديــرات الصناعة إلى أن مبيعات كاوبويز تشــكِّل ٢٠٪ من جميع مبيعات البضائع في اتحاد كرة القدم الأمريكي؛ مما يعكس حقيقة أن كاوبويز هو صاحب أكثر حقوق الامتياز الرياضية شهرةً في العالم.

وفقًا لإريك لاي، مراســل موقع ComputerWorld؛ فقــد يكون توني رومو وبقية فريق دالاس كاوبويز متواضعي المســتوى فقط في مجال كرة القدم في السنوات القليلة الماضية، ولكن بعيدًا عن الملعب، وخاصةً في ميدان التجارة؛ فإنهم يظلون فريقاً أمريكيًّا.

أسئلة للمناقشة:

١- كيف استخدم دالاس كاوبويز تصوير المعلومات؟

٢- ما هو التحَدِّي والحل المقترح والنتائج المتحصَّل عليها؟

Sources: Lai, E. (2009, October 8). BI visualization tool helps Dallas Cowboys sell more Tony Romo jerseys. ComputerWorld; Tableau case study. tableausoftware.com/learn/stories/tableau-andteknion-exceed-cowboys-requirements (accessed July 2016).

تصميم لوحة المعلومات:

لا تُعَدُّ لوحات المعلومات مفهومًا جديدًا. فيمكن أن تعود جذورها التاريخية على الأقل إلى نظام المعلومات التنفيذية في ثمانينيات القرن العشرين. واليوم، تتواجد لوحات المعلومات في كلِّ مكان. فعلى سبيل المثال: قبل عِدة سنوات، قدَّرت Forrester Research أن أكثر من ٤٠٪ من أكبر ٢٠٠٠ شركة في العالم استخدمت هذه التقنية (أنتي آند مكجريجور، ٢٠٠٦). ومنذ ذلك الحين؛ يمكننا أن نعتبر بمنتهى الثقة أن هذا الرقم قد ارتفع بشكلٍ كبير. وفي الواقع؛ في هذه الأيام سيكون من غير المعتاد رؤية شركة كبيرة تستخدم نظامًا لذكاء الأعمال لا يعتمد على لوحات التحكم في الأداء. كما يُقدِّم موقع Dashboard Spy (dashboardspy.com/about) دليلًا آخر على وجودها في كلِّ مكان. ويحتوي الموقع على أوصاف ولقطات شاشة لآلاف من لوحات المعلومات ذكاء الأعمال، وبطاقات النتائج، وواجهات ذكاء الأعمال المُستَخدَمة من قِبل الشركات من جميع الأحجام والصناعات، والمنظمات غير الرِّبحية، والوكالات الحكومية.

وفقاً لـ Eckerson (2006)، وهو خبيرٌ معروفٌ في ذكاء الأعمال بشكلٍ عام ولوحات المعلومات على وجه الخصوص؛ فإن الخاصية الأكثر تميُّزًا في لوحة المعلومات هي تمتُّعها بثلاثِ طبقات من المعلومات:

١- **المراقبة:** البيانات الرسومية المُستَخرَجة لرَصْد مقاييس الأداء الرئيسية.

٢- **التحليل:** تلخيص البيانات ذات الأبعاد لتحليل السبب الجذري للمشكلات.

٣- **الإدارة:** تفصيل البيانات التشغيلية التي تحَدِّد الإجراءات التي يجب اتخاذها لحلِّ مشكلةٍ ما.

بسبب هذه الطبقات؛ فإن لوحات المعلومات تحوي الكثير من المعلومات في شاشة واحدة. ووفقاً لـ Few (2005)، «يتمثل التحَدِّي الأساسي لتصميم لوحة المعلومات في عرض جميع المعلومات المطلوبة على شاشة واحدة بوضوح وبدون تشتيت، بطريقة يمكن استيعابها بسرعة». ولسرعة

استيعاب الأرقام؛ يلزم وَضْع الأرقام في السياق. ويُمكِن القيام بذلك عن طريق مقارنة الأرقام محل الاهتمام بالأرقام الأخرى الأساسية أو المُستَهدَفة؛ وذلك بالإشارة إلى ما إذا كانت الأرقام جيدةً أم سيئةً، عن طريق معرفة ما إذا كان الاتجاه أفضل أم أسوأ، وباستخدام أدوات أو عناصر عرض متخصصة لوضع السياق النسبي والتقديري. وتشتمل بعض المقارنات الشائعة التي يتمُّ إجراؤها عادةً في أنظمة ذكاء الأعمال على مقارنات مع القِيم السابقة والمتوقعة والمُستَهدَفة والقياسية أو القيم المتوسطة والمتعددة لنفس القياس، وقِيَم المقاييس الأخرى (مثل: الإيرادات مقابل التكاليف). وحتى مع القياسات النسبية، من المهم أن نشير على وجه التحديد ما إذا كان رقمٌ معينٌ جيدًا أم سيئًا وما إذا كان متجهًا في الاتجاه الصحيح أم لا. وبدون هذه الأنواع من الدِّلالات التقييمية، قد يستغرق الأمرُ وقتًا طويلًا لتحديد حالة رقمٍ ما أو نتيجةٍ معينة. وعادةً؛ يتمُّ استخدامُ أيِّ كائنات مرئية متخصصة (مثل: إشارات المرور، والمفاتيح، والمعايير) أو السِّمات المرئية (مثل: ترميز الألوان) لتعيين السياق التقديري. ويبرز في الحالة العملية ٢-٨ حلٌّ تفاعليٌّ لاستكشاف بيانات التقارير، يقوم هذا الحلُّ على لوحة المعلومات، وقد تمَّ بناؤه بواسطة شركة طاقة.

حالة عملية ٢-٨

تساعد التحليلات المرئية مورد الطاقة في عمل روابط أفضل

تشهدُ أسواق الطاقة في جميع أنحاء العالم تغيُّرًا وتحوُّلًا كبيرين؛ مما يصنع فرصًا كبيرةً إلى جانب تحَدِّيات كبيرة. وكما هو الحال في أيِّ صناعة، تجتذب الفرصُ عددًا أكبر من اللاعبين في السوق؛ مما يزيد المنافسة، ويُقلِّل من مستوى التفاوت في اتخاذ القرارات التجارية بشكلٍ أقل من المثالية. ويتطلَّب النجاح إنشاء ونشر معلومات دقيقة وفي الوقت المناسب وتقديمها إلى أيِّ شخص، وكلما دعت الحاجة. فعلى سبيل المثال: إذا كنتَ بحاجةٍ إلى تتبُّع كلٍّ من ميزانيات التسويق بسهولة، والتوازن بين أحمال عمل الموظفين، واستهداف العملاء برسائل تسويقية مُخصَّصة؛ فستحتاج إلى ثلاثة حلول مختلفة لإعداد التقارير. تقوم شركة Electrabel GDF SUEZ بكلِّ ذلك؛ من أجل وحدة أعمال التسويق والمبيعات الخاصة بها مع مِنصَّة تحليلات SAS المرئية.

إن نهج الحلِّ الواحد هو توفيرٌ كبيرٌ للوقت لمحترفي التسويق في صناعةٍ تشهدُ تغيراً هائلاً. يقول داني نوب -مدير إعداد التقارير والتطوير في وحدة أعمال التسويق والمبيعات في Electrabel: «إنه تَحَدٍّ كبير لاستقرار مركزنا السُّوقي في سوق الطاقة. ويشملُ ذلك الحجمَ والأسعارَ وهوامشَ الرِّبح لكلٍّ من عملاء التجزئة والأعمال التجارية». وتُعَدُّ هذه

الشركــة أكبرَ موردٍ للكهرباء في بلجيكا وأكبر منتــجٍ للكهرباء في كلٍّ من بلجيكا وهولندا. يقــول نوب: «إنه من الأهمية بمــكان أن تزيد Electrabel من كفاءة اتصالات عملائها؛ إذ إنها تستكشف القنوات الرقمية الجديدة وتطوِّر خدمات جديدة متعلقة بالطاقة».

ويضيف نوب: «كلما عرفنا العملاء بشــكلٍ أفضل؛ كان احتمالُ نجاحنا أكبر، وهذا هو السببُ في أننا نجمع المعلومات من مصادر متنوعة، مثل: حركة الهاتف مع العميل، والأسئلة عبر الإنترنت، والرسائل النصية، وحملات البريد. وستُشكِّل هذه المعرفة المُعزَّزة لعملائنا وقاعدة العملاء المحتملين مَيْزةً إضافيةً في سوقنا التنافسية».

نسخة واحدة من الحقيقة:

كانت Electrabel تســتخدم مِنصات وأدوات متنوعةً لأغــراض إعداد التقارير. وقد أدَّى ذلك أحيانًا إلى غموضٍ في الأرقام المذكورة. وقد واجهت المؤسســة أيضًا مشــكلاتٍ في الأداء في معالجــة وحدات مســتودعات البيانات الكبيرة. تزيــل تحليلات SAS المرئية مع تقنية الذاكرة الغموض ومشكلات الأداء. يقول نوب: «لدينا الاستقلالية والمرونة للاستجابة للحاجــة إلى رؤيــة العملاء وتصوير البيانات داخليًا. وبعد كلِّ هــذا؛ يُعَدُّ الإعدادُ السريع للتقارير أحد المتطلبات الأساسية للإدارات الموجَّهة عمليًا مثل المبيعات والتسويق».

العمل بشكلٍ أكثر كفاءة، وبتكلفةٍ أقل:

تســعى تحليــلاتُ SAS المرئية إلى تشــغيل عملية تحديث المعلومــات في التقارير بصورةٍ آلية. فبدلًا من إنشاء تقرير قديم في وقت اكتماله؛ يتمُّ تحديث البيانات لجميع التقاريــر مرةً واحدةً في الأســبوع وإتاحتها عــلى لوحات المعلومــات. وفي نشر الحل، اختارت Electrabel نهجًا تدريجيًا يبدأ بتقارير بسيطة وينتقل إلى تقارير أكثر تعقيدًا. وقد احتاج التقريرُ الأول بضعة أســابيع لبنائه، ثم جاءت بقية التقارير بشــكلٍ أسرع. وتتضمَّن النجاحات ما يلي:

- تستغرق البياناتُ التي استغرقت يومين للتحضير الآن ساعتين فقط.

- مَسْح رؤية بيانية لفواتير وتكوين الفواتير للعملاء B2B.

- تقرير إدارة عبء العمل من قِبل الفِرق التشغيلية. مما يمكِّن المديرين من تقييم أعباء فريق العمل على أساس أسبوعي أو طويل الأجل ويمكِّنهم من إجراء تعديلات وفقًا لذلك.

يقول نوب: «لقد قمنا بتحســين كفاءة أعمالنا بشــكلٍ كبير، ويمكننا تقديمُ بيانات

عالية الجودة، وتقارير بشكلٍ أكثر تكرارًا وبتكلفة أقلّ بكثير. وإذا كانت الشركة بحاجةٍ إلى تجميع البيانات من مصادر متعددة؛ فهذا أيضًا من السهولة بمكان. كما يمكن إنشاءُ تقارير مرئية، استنادًا إلى هذه البيانات، خلال بضعة أيام، أو حتى بضع ساعات».

ويضيف: «إنَّ الشركة تُخطِّط لمواصلة توسيع رؤيتها في السلوك الرقمي لعملائها، والجمع بين البيانات من تحليلات الويب والبريد الإلكتروني والوسائل الإعلامية للتواصل الاجتماعي مع البيانات من أنظمة النهاية الخلفية. وفي النهاية؛ نحن نريدُ استبدال جميع التقارير التي تحتاج إلى عمل مكثف بتحليلات SAS المرئية». وكما يقول نوب مضيفًا: إن مرونة تحليلات SAS المرئية أمرٌ بالغ الأهمية لإدارته: «وهذا سيتيح لنا المزيدَ من الوقت للتصَدِّي للتحَدِّيات الأخرى. ونريد أيضًا أن نجعل هذه الأداة متاحةً على أجهزتنا المحمولة. وسيسمح ذلك لمديري الحسابات لدينا باستخدام التقارير الحديثة والثاقبة والقابلة للتكيُّف عند زيارة العملاء. كما يوجد لدينا مِنصَّة لإعداد التقارير الموجهة نحو المستقبل للقيام بكلِّ ما نحتاجه».

أسئلة للمناقشة:

١- لماذا تعتقدُ أن شركات إمدادات الطاقة من بين المستخدمين الرئيسيين لأدوات تصوير المعلومات؟

٢- كيف استخدم Electrabel تصوير المعلومات للنسخة الواحدة من الحقيقة؟

٣- ما التحديات التي واجهوها؟ وماذا كان الحل المقترح؟ وما النتائج التي تمَّ الحصولُ عليها؟

Source: SAS Customer Story, “Visual analytics helps energy supplier make better connections” at http://www.sas.com/en_us/cus-tomers/electrabel-be.html (accessed July 2016).

ما الذي تبحث عنه في لوحة المعلومات:

على الرغم من اختلاف أداء لوحات المعلومات والأُطُر الأخرى لتصوير المعلومات؛ فإنها جميعها تشترِكُ في بعض الخصائص الشائعة للتصاميم. **أولًا:** إنها جميعًا ملائمة لنظام ذكاء الأعمال الأكبر أو نظام قياس الأداء. وهذا يعني أنَّ البنية الأساسية الخاصة بهم جميعًا هي ذكاء

الأعـمال أو بنيـة إدارة الأداء في النظام الأكبر. **ثانيًا:** تتمتَّع كلُّ لوحات المعلومات المُصمَّمة جيدًا وتصويرات المعلومات الأخرى بالخصائص التالية (Novell, 2009):

- جميعها تستخدم مكوِّنات بصرية (مثل: الرسومات البيانية، وأشرطة الأداء، وخطوط المؤشرات، والمقاييس، والأمتار، وإشارات المرور الضوئية)؛ لتسليط الضوء - في لمحة واحدة - على البيانات والاستثناءات التي تتطلَّب اتخاذ إجراء.
- تُعَدُّ جميعها شفافةً للمستخدِم، وهذا يعني أنها تتطلَّب الحَدَّ الأدنى من التدريب وتكون سهلة الاستخدام للغاية.
- جميعها تجمِّع البيانات من مجموعةٍ متنوعةٍ من الأنظمة إلى عَرْض واحدٍ وموجزٍ ومُوحَّدٍ للأعمال.
- إنها جميعًا تمكِّن من التمحيص أو الانتقال إلى مصادر البيانات الأساسـية أو التقارير؛ مما يوفر مزيداً من التفاصيل حول السياق الأساسي النِّسبي والتقديري.
- تُقـدِّم جميعها عرضًا حيويًّا للعالم الحقيقي مع تحديث البيانات في الوقت المناسـب؛ مما يتيح للمُستخدِم النهائي مواكبة أحدث التغييرات في النشاط التجاري.
- تتطلَّب جميعها القليل من الترميز المُخصَّص - إن وجد - لتنفيذه ونشره وصيانته.

أفضل الممارسات في تصميم لوحة المعلومات:

إن ترديد العقارات لمقولة «الموقع، الموقع، الموقع» تُظهِر بوضوح أن السِّـمة الأكثر أهميةً لجزء من الملكية العقارية هي المكان الذي تُوجَد فيه. أمَّا بالنسبة إلى لوحات المعلومات فالترديدُ يكون «البيانات، البيانات، البيانات». فغالبًا ما تكون البيانات؛ هي أحد أهم الأشياء التي يجبُ أخذها في الاعتبـار عند تصميم لوحات المعلومات (Carotenuto, 2007). حتى إذا كان مظهرُ لوحة القيادة يبدو احترافيًّا، وممتعاً من الناحية الجمالية، ويتضمَّن رسومًا بيانية وجداول تمَّ إنشاؤها وفقًا لمعايير التصميـم المرئي المقبولة؛ فمن المهمِّ أيضًا أن تسـأل عن البيانات: هـل هي موثوقة؟ هل هي في الوقت المناسب؟ هل هناك أيُّ بيانات مفقودة؟ هل هي ثابتةٌ عبر جميع لوحات المعلومات؟ فيما يلي بعضٌ من أفضل الممارسات القائمة على الخبرة في تصميم لوحة المعلومات (Radha, 2008).

مؤشرات الأداء الرئيسية مع معايير الصناعة:

يرغـب العديدُ مـن العملاء في وقتٍ ما في معرفة ما إذا كانت المقاييس التي يقيسـونها هي المقاييـس الصحيحة التي يجبُ مراقبتها. وفي بعض الأحيان؛ يكتشـف العملاءُ أن المقاييس التي

يتتبعونها ليست هي المقاييس اللازم تتبُّعها. إن إجراء تقييمٍ للثغرات مع معايير الصناعة يجعلك تتوافق مع أفضل الممارسات في المجال.

لفُّ مقاييس لوحة المعلومات مع البيانات الوصفية السياقية:

في كثيرٍ من الأحيان عندما يتمُّ تقديم تقريرٍ أو لوحة معلومات/ بطاقة نتائج مرئية لمستخدمي الشركات؛ تظلُّ هناك أسئلةٌ بدون إجابة. وفيما يلي بعض الأمثلة على ذلك:

- ما مصدر هذه البيانات؟
- في أثناء تنزيل مستودع البيانات، ما النسبة المئوية للبيانات التي تمَّ رفضها أو واجهت مشكلات جودة البيانات؟
- هل تُقدِّم لوحة المعلومات معلومات «جديدة» أم معلومات «قديمة»؟
- متى كانت آخر مرة تمَّ فيها تحديثُ مستودع البيانات؟
- ومتى سيتمُّ تحديثه لاحقًا؟
- هل كان من شأن أيِّ صفقة عالية القيمة تحريفُ الاتجاهات العامة المرفوضة كجزءٍ من عملية التنزيل؟

التحقُّق من صحة تصميم لوحة المعلومات بواسطة أخصائي القابلية للاستخدام:

في معظم بيئاتِ لوحة المعلومات؛ يتمُّ تصميمُ لوحة المعلومات بواسطة أخصائي أدوات دون الأخذ في الاعتبار بمبادئ قابلية الاستخدام. وعلى الرغم من أن مستودع البيانات يكون جيدَ التصميم وقادراً على أن يُحقِّق أداءً جيدًا؛ فإنَّ العديدَ من مُستخدِمي الأنشطة التجارية لا يستخدمون لوحة المعلومات؛ نظراً لعدم مواءمتها مع المستخدمين (وفقاً لاعتقادهم السائد)؛ مما يؤدي إلى سوء اتخاذ بنية أساسية وإلى مشكلات في إدارة التغيير. ويمكن أن يؤدي التحقق المُسبَق من تصميم لوحة المعلومات بواسطة أخصائي قابلية الاستخدام إلى تخفيف هذه المخاطرة.

إعطاء أولويات وترتيب التنبيهات/ الاستثناءات المتدفقة إلى لوحة المعلومات:

بسبب وجود أطنان من البيانات الخام؛ فمن الأهمية بمكان وجودُ آلية لدفع الاستثناءات/ السلوكيات المهمة بشكلٍ استباقي إلى مستهلكي المعلومات. ومن الممكن تقنينُ قاعدة الأعمال التي تضبط نمط التنبيه محلَّ الاهتمام. كما يمكِن أن يتمَّ ترميزها في برنامج، باستخدام الإجراءات المُخزَّنة في قواعد البيانات، والتي يمكنها الظهورُ من خلال جداول الحقائق واكتشاف الأنماط

التــي تحتاج إلى انتباه فوري. وبهذه الطريقة، تعثر المعلومات على مســتخدم الأعمال في مقابل مستخدم الأعمال الذي يتفحص جداول الحقائق لظهور الأنماط الحرجة.

إثراء لوحة المعلومات بتعليقات مستخدِم الأعمال:

عندمــا يتمُّ تقديمُ نفس لوحة المعلومات إلى عِدة مســتخدمين؛ فيمكن توفيرُ مربع نص صغير يمكنه التقاط التعليقات من وجهة نظر أيِّ مُستَخدمٍ نهائي. وبهذا يمكِن غالبًا وَضْع علامة على لوحة المعلومات لوضع المعلومات في سياقها؛ مع إضافة وجهة النظر إلى KPIs المهيكلة التي يتمُّ تقديمها.

تقديم المعلومات في ثلاثة مستويات مختلفة:

يُمكِن تقديم المعلومات في ثلاث طبقات حســب تقسيمات المعلومات: مستوى لوحة المعلومات المرئية، ومستوى التقرير الثابت، ومستوى مكعب الخدمة الذاتية. وعندما يتنقل المستخدِم في لوحة المعلومــات، يمكن تقديم مجموعة بســيطة من ٨ إلى ١٢ KPIs؛ مما يعكــس انطباعًا حول ما يجري بشكلٍ جيد وما هو غير جيد.

اختيار البنية المرئية الصحيحة باستخدام مبادئ تصميم لوحة المعلومات:

في عَرْض المعلومات على لوحة المعلومات؛ يتمُّ تقديمُ بعض المعلومات بشكلٍ أفضل باستخدام شريط الرسوم البيانية، وبعضها يكون أفضل مع المخططات البيانية للسلاسل الزمنية، وعند عَرْض الارتباطــات، يكون مربع التبعثر هو الأفضل. وفي بعضِ الأحيان مجرد تقديم المعلومات كجداول بســيطة يكون أكثر فاعلية. وبمجرد أن يتمَّ توثيقُ مبادئ تصميم لوحة المعلومات بشكلٍ صريح، يمكن لجميع المطوِّرين الذين يعملون على الواجهة الأمامية الالتزام بنفس المبادئ في أثناء تقديم التقارير ولوحة المعلومات.

توفير التحليلات الإرشادية:

في منظمةٍ نموذجية؛ يمكِن للمستخدمين من رجال الأعمال أن يكونوا على مستويات مختلفة من النضج التحليلي. كما يمكِن اســتخدامُ إمكانية لوحة المعلومات لتوجيه مستخدِم الأعمال «متوسط المستوى» إلى الوصول إلى نفس مسار التنقل الذي يسلكه مستخدمُ الأعمال المُحنَّك بطريقة تحليلية.

أسئلة مراجعة على القسم ٢-١١:

١- ما هي لوحة المعلومات؟ ولماذا هي منتشرة جداً؟

٢- ما هي عناصرُ الرسوم البيانية المُستَخدَمة عادةً في لوحات المعلومات؟ ولماذا؟

٣- اذكر مع الشرح الطبقات الثلاث من المعلومات المصوَّرة على لوحات المعلومات.

٤- ما هي الخصائص المشتركة للوحات المعلومات وغيرها من المعلومات المرئية؟

٥- ما هي أفضل الممارسات في تصميم لوحة القيادة؟

تلخيص لأهمِّ نقاط الفصل:

- أصبحت البياناتُ واحدةً من أكثر الأصول قيمةً لمنظمات اليوم.
- تُعَدُّ البيانات هي العنصر الرئيسي لأيِّ مبادرة ذكاء أعمال، وعلم البيانات، ومبادرة تحليلات الأعمال.
- على الرَّغم من أن القيمة المقترحة لها لا يمكن إنكارها، وللوفاء بوعودها، يجبُ أن تلتزم البيانات ببعض المقاييس الأساسية للاستخدام والجودة.
- تشيرُ البيانات (المعطاة في صورة مفردة) إلى مجموعةٍ من الحقائق التي يتمُّ الحصولُ عليها عادةً كنتيجة للتجارب أو المشاهدات أو المعاملات أو الخبرات.
- عند أعلى مستويات الغموض؛ يُمكِنُ تصنيفُ البيانات على أنها مهيكلةٌ وغير مهيكلة.
- لا تكون البيانات في شكلها الأصلي/ الخام جاهزةً عادةً لتكون مفيدةً في مهامِّ التحليلات.
- تُعَدُّ المعالجة المُسبَقة للبيانات مهمةً مُملَّةً ومتطلبةً للوقت؛ ولكنها ضروريةٌ في تحليلات الأعمال.
- يُعرَّف الإحصاءُ على أنه مجموعةٌ من التقنيات الرياضية لتوصيف وتفسير البيانات.
- يُمكِن تصنيف الأساليب الإحصائية على أنها إمَّا وصفية أو استنتاجية.
- تُعَدُّ الإحصائيات بشكلٍ عام والإحصائيات الوصفية بشكلٍ خاص جزءًا مهمًّا من تحليلات ذكاء الأعمال وتحليلات الأعمال.
- يُمكِن استخدامُ طرق الإحصاء الوصفي لقياس النزعة المركزية، أو التشتُّت، أو شكل مجموعة بيانات معينة.
- قد يكون الانحدارُ، وخاصةً الانحدار الخَطِّي، هو أكثرُ تقنيات التحليلات شهرةً واستخدامًا في الإحصاءات.
- يُنظَر لكلٍّ من الانحدار الخطي والانحدار اللوجستي على أنهما النوعان الرئيسان من الانحدارات في الإحصاءات.
- يُعرَف الانحدار اللوجستي كخوارزمية تصنيف قائمة على الاحتمالية.

- تُعرَّف السلسـلة الزمنية على أنها سلسـلة من نقاط البيانات للمتغير، يتمُّ قياسها وتسجيلها في نقاط متتالية من الزمن متباعدة على فترات زمنية مُوحَّدة.
- يُعَدُّ التقريرُ أيَّ عملية تواصل من صُنع الإنسـان تمَّ تحضيرها بغرضٍ معين؛ لتوصيل المعلومات بصيغة قابلة للتقديم أو العرض.
- يُعَدُّ تقريرُ الأعمال مستنداً مكتوباً يحتوي على معلومات تتعَلَّق بأمور الأعمال.
- إنَّ مفاتيح أيِّ تقرير تجاري ناجحٍ؛ هي: الوضوحُ، والإيجازُ، والاكتمالُ، والصحة.
- يُعَدُّ التصويرُ المرئي للبيانات؛ هو استخدامُ التمثيلات المرئية؛ لاستكشاف البيانات وفهمها وتوصيلها.
- قد يكون المخطط البياني الأبرز للمعلومات في الماضي؛ هو ذلك الذي طوَّره Charles J. Minard، والذي صوَّر بيانيًّا الخسائر التي تكبَّدها جيش نابليون في الحملة الروسية عام ١٨١٢م.
- تتضَمَّن أنواعُ الرسـوم البيانية الأساسـية الرَّسـم البياني الخَطِّي، وشريط الرسـم البياني، والرسم البياني الدائري.
- غالبًا ما يتمُّ اشتقاق الرسوم البيانية المتخصصة من المخططات الأساسية كحالات استثنائية.
- إن تقنيات وأدوات عرض البيانات تجعل من مستخدمي تحليلات الأعمال وأنظمة ذكاء الأعمال أفضل مستهلكين للمعلومات.
- تُعرَّف التحليلات المرئية بأنها مزيجٌ من التصوير والتحليلات التنبؤية.
- أدَّى ازديـادُ الطلـب على التحليلات البصرية إلى جانب حجم البيانات سريع النمو إلى نموٍّ هائلٍ في استثمارات أنظمة التصوير ذات الكفاءة العالية.
- توفر لوحات المعلومات عروضاً مرئيةً لمعلومات مهمة يتمُّ دَمْجُها وترتيبها على شاشة واحدة؛ حتى يمكن استيعاب تلك المعلومات في لمحةٍ واحدةٍ والتعمُّق فيها بسهولة واستكشافها بشكلٍ أكبر.

مصطلحات أساسية:

التحليلات جاهزة	تصوير البيانات	الانحدار الخطي	بيانات النسبة
الوسط الحسابي	معطيات	الانحدار اللوجيستي	الانحدار
box-and-whiskers plot	الإحصائيات الوصفية	متوسط الانحراف المطلق	التقرير
box plot	تخفيض الأبعاد	الوسيط	مربع التبعثر

التحليلات جاهزة	تصوير البيانات	الانحدار الخطي	بيانات النسبة
الرسم البياني الفقاعي	التشتُّت	المنوال	الانحراف
تقرير الأعمال	عالي الأداء	بيانات اسمية	الانحراف المعياري
بيانات فئوية	الحوسبة	تحليلات على الإنترنت	الإحصاء
المركزية	المدرَّج التكراري	المعالجة (OLAP)	السَّرد القصصي
الارتباط	الإحصاء الاستنتاجي	بيانات ترتيبية	بيانات مهيكلة
لوحات المعلومات	الأداء الرئيسي	المربعات الصغرى العادية	التنبؤ بالسلاسل الزمنية
المعالجة المسبقة للبيانات	مؤشر (KPI)	(OLS)	بيانات غير مهيكلة
جودة البيانات	المعرفة	الرسم البياني الدائري	اختيار المتغير
أمن البيانات	التفرطح	الرُّبيع	التباين
تصنيف البيانات	التعلُّم	النطاق	التحليلات المرئية

أسئلة للمناقشة:

١- كيف تَصِف أهمية البيانات في التحليلات؟ هل يمكننا التفكيرُ في التحليلات بدون بيانات؟ اشرح.

٢- بالنظر إلى التعريف الجديد والواسع لتحليلات الأعمال؛ ما هي المدخلات والمخرجات الرئيسية لاستمرارية التحليلات؟

٣- من أين تأتي بيانات تحليلات النشاط التجاري؟ وما هي مصادرُ وطبيعة تلك البيانات الواردة؟

٤- ما هي المقاييس الأكثر استخدامًا في البيانات الجاهزة للتحليلات؟

٥- ما هي الفئات الرئيسة للبيانات؟ وما هي أنواعُ البيانات التي يمكن استخدامها في ذكاء الأعمال والتحليلات؟

٦- هل يُمكِننا استخدامُ نفس تمثيل البيانات لجميع نماذج التحليلات (بمعنى: هل تتطلَّب نماذج التحليلات المختلفة مخططًا مختلفًا لتمثيل البيانات)؟ ولماذا؟ أو لماذا لا؟

٧- لماذا لا تكون البيانات الأصلية/ الأوَّلية قابلةً للاستخدام بسهولة بواسطة مهامِّ التحليلات؟

٨- ما هي الخطوات الرئيسة لمعالجة البيانات؟ اذكر مع الشرح أهميتها في التحليلات.

٩- ماذا يعني تنظيف/ فرك البيانات؟ وما هي الأنشطة التي يتمُّ أداؤها في هذه المرحلة؟

١٠- يُمكِــن تطبيق تخفيض البيانات على الصفوف (أخذ العينات) و/ أو الأعمدة (اختيار متغير). أيُّهما أكثر تحديًّا؟ اشرح.

١١- مــا هي العلاقة بين الإحصائيات وتحليــلات الأعمال (مع الأخذ في الاعتبار مكانة الإحصاءات من تصنيف تحليل الأعمال)؟

١٢- ما هي الاختلافات الرئيسة بين الإحصاء الوصفي والاستقرائي؟

١٣- ما هو a box-and-whiskers plot؟ وما هي أنواع المعلومات الإحصائية التي تمثلها؟

١٤- ما هي خصائصُ الشكل الأكثر شيوعًا لوصف توزيع البيانات؟

١٥- اذكر مقاييس النَّزعة المركزية للإحصاءات الوصفية، مع تعريف موجز لكلِّ مقياس.

١٦- ما هي أوجه الشبه والاختلاف بين الانحدار والارتباط؟

١٧- اذكر مع الشرح الخطوات الرئيسة لمتابعة التطوُّر في نموذج الانحدار الخطي.

١٨- مــا هي أكثر الافتراضات الواضحة شــيوعًا للانحدار الخطي؟ وما الذي يلزم لنماذج الانحدار في مقابل هذه الافتراضات؟

١٩- ما هي أوجه الشبه والاختلاف بين الانحدار الخطي والانحدار اللوجستي؟

٢٠- ما هي السَّلاسل الزمنية؟ وما هي تقنيات التنبؤ الرئيسية لبيانات السلاسل الزمنية؟

٢١- ما هو تقريرُ العمل؟ ولماذا يتمُّ الاحتياجُ إليه؟

٢٢- ما هي أفضل الممارسات في إعداد تقارير الأعمال؟ وكيف يمكننا جَعْل تقاريرنا تتميَّز عن غيرها؟

٢٣- اشرح عملية دورة الإدارة مع التعليق على دور تقارير الأعمال.

٢٤- اذكر مع الشرح الفئات الرئيسية الثلاثة لتقارير الأعمال.

٢٥- لمــاذا أصبــح تصويرُ المعلومات محوريًّا في تحليلات ذكاء الأعــمال وتحليلات الأعمال؟ وهل هناك فرقٌ بين تصوير المعلومات والتحليلات المرئية؟

٢٦- ما هي الأنواع الرئيسة من الرسوم/ المخططات البيانية؟ ولماذا يوجد الكثير منها؟

٢٧- كيف يمكنك تحديدُ الرسم البياني المناسب لكلِّ مهمة؟ اشرح وبَرِّر إجابتك.

٢٨- ما هو الفرق بين تصوير المعلومات والتحليلات المرئية؟

٢٩- لماذا يجبُ أن يكون السرد القصصي جزءًا من إعدادك للتقارير وتصويرك للبيانات؟

٣٠- ما هي لوحات المعلومات؟ وماذا تعرض؟

٣١- ما هي أفضل الممارسات في تصميم لوحات المعلومات الغنية بالمعلومات؟

٣٢- هل تعتقدُ أن لوحات المعلومات/ الأداء موجودة لتبقى؟ أم هي على وشك التقادم؟ وبرأيك ماذا ستكون الموجة الكبيرة التالية في ذكاء الأعمال وتحليلات الأعمال؛ من حيث تصوير البيانات/ المعلومات؟

تمارين

جامعة تيراداتا وغيرها من التمارين اليدوية الأخرى

١- قم بتنزيل بيانات «سلوك التصويت» ووصف البيانات المختصر من موقع الكتاب على الإنترنت. هذه مجموعة بيانات تمَّ تجميعها يدويًّا من المقاطعات في جميع أنحاء الولايات المتحدة. تتمُّ معالجة البيانات جزئيًّا، أيْ: يتمُّ إنشاء بعض المتغيرات المشتقة. وتتمثل مهمتك في إجراء معالجة مسبقة للبيانات من خلال تحديد الخطأ والشذوذ، واقتراح الإصلاحات والحلول. وفي النهاية؛ يجب أن يكون لديك إصدارٌ جاهزٌ للتحليلات لهذه البيانات. وبمجرد اكتمال المعالجة المسبقة، اسحب هذه البيانات إلى Tableau (أو إلى أداة برمجية أخرى لتصوير البيانات)؛ لاستخراج معلومات مرئية مفيدة منها. وللقيام بذلك؛ قم بوَضْع الأسئلة والفرضيات ذات الصِّلة (ابتكر ثلاثةً منها على الأقل)، وقم بإنشاء تصويرات ملائمة تتعامل مع تلك الأسئلة لـ «اختبارات» تلك الفرضيات.

٢- قم بتنزيل Tableau (من على tableau.com، باتباع الإرشادات الأكاديمية لتنزيل البرامج المجانية من على موقعها). باستخدام مجموعة البيانات Visualization_MFG_Sample (المتاحة على ملف Excel على موقع هذا الكتاب)، أجب عن الأسئلة التالية:

١- ما هي العلاقة بين إجمالي إيرادات شباك التذاكر والمعلمات الأخرى المتعلِّقة بالفيلم الواردة في مجموعة البيانات؟

٢- كيف تختلف هذه العلاقة عبر سنوات مختلفة؟ إعداد تقرير مكتوب ذي مظهر احترافي يتمُّ تحسينه باستخدام لقطات من نتائجك الرسومية.

٣- انتقل إلى teradatauniversitynetwork.com. ابحث عن مقالة تتناول طبيعة البيانات، وإدارة البيانات، و/ أو تقنين البيانات من حيث صِلتها بذكاء الأعمال والتحليلات، ثم حلِّل محتوى المقالة بشكلٍ نقدي.

٤- انتقل إلى مستودع بيانات UCI (archive.ics.uci.edu/ml/datasets.html)، وحَدِّد مجموعةً كبيرةً من البيانات تحتوي على كلٍّ من القِيم الرقمية والقِيم الاسمية. باستخدام Microsoft Excel أو أيّ برنامج إحصائي آخر:

١- قم بحساب وتفسير مقاييس النَّزعة المركزية لكلٍّ متغير.

٢- قم بحساب وتفسير مقاييس التشتُّت/ الانتشار لكلٍّ متغير.

٥- انتقل إلى مستودع بيانات UCI (archive.ics.uci.edu/ml/datasets.html)، وحَدِّد مجموعتين من البيانات، إحداهما لـ التقدير/ الانحدار والأخرى للتصنيف. باستخدام Microsoft Excel أو أيّ برنامج إحصائي آخر:

١- قم بتطوير وتفسير نموذج انحدار خطِّي.

٢- قم بتطوير وتفسير نموذج انحدار لوجستي.

٦- انتقل إلى موقع KDnuggest.com، وتعرَّف على مجموعة موارد التحليلات المتاحة على هذه البوابة. ثم حَدِّد مقالةً أو ورقةً بيضاء، أو نصًّا للمقابلة يتناول طبيعة البيانات، و/ أو إدارة البيانات، و/ أو تقنين البيانات؛ من حيث صِلتها بذكاء الأعمال وتحليلات الأعمال، ثم حلِّل محتوى المقالة بشكلٍ نقدي.

٧- انتقل إلى مُدوَّنة ستيفن فيو، [("The Perceptual Edge" (perceptualedge.com]. انتقل إلى قسم "أمثلة". اقرأ بعضًا من هذه الأمثلة. اذهب الآن إلى dundas.com، حَدِّد قسم "Gallery" في الموقع. بعد ذلك، انقر على "Digital Dashboard". سيعرض لك مجموعةً متنوعةً من العروض التوضيحية المختلفة للوحة المعلومات. قم بتشغيل اثنين من هذه العروض التوضيحية.

١- ما هي أنواع المعلومات والمقاييس التي تظهر في العروض التوضيحية؟ وما هي أنواع الإجراءات التي يمكنك اتخاذها؟

٢- باستخدام بعض المفاهيم الأساسية من نقد Few، قُمْ بوصف بعض نقاط التصميم الجيدة ونقاط التصميم السيئة للعروض التوضيحية.

٨- قم بتنزيل أداة لتصوير المعلومات، مثل Tableau أو QlikView أو Spotfire. إذا لم يكن لدى مدرستك اتفاقية تعليمية مع هذه الشركات؛ فستكون النسخة التجريبية كافية لهذا التمرين. استخدم بياناتك الخاصة (إذا كان لديك أيٌّ منها) أو استخدم إحدى مجموعات البيانات التي تأتي مع الأداة (وعادةً ما تحتوي على مجموعةِ بيانات واحدة أو أكثر لأغراض العرض

التوضيحي). قم بدراسة البيانات، وافترض مشكلتين من ابتكارك من مشكلات الأعمال، واستخدم التصوير البياني للبيانات لتحليل وعرض هاتين المشكلتين، ثم اقترح حلًا محتملًا لهما.

٩- انتقل إلى: teradatauniversitynetwork.com. ابحث عن "مشروع برنامج Tableau". اقرأ الوصف، نفِّذ المهامَّ، ثم أجب عن الأسئلة.

١٠- انتقل إلى: teradatauniversitynetwork.com. ابحث عن مهامِّ تحليلات SAS المرئية. باستخدام المعلومات والإرشادات من نوع خطوة بخطوة المتوفرة في المهمة؛ قم بتنفيذ التحليل على أداة SAS Visual Analytics (وهو نظامٌ يدعم الويب ولا يتطلب أيَّ تثبيتٍ محلي). أجب عن الأسئلة المطروحة في المهمة.

١١- ابحث عن مقالتين على الأقل (إحداهما مقالٌ صحفي والآخر ورقةٌ بيضاء) تتحدثان عن سَرْد القصص، وخصوصًا في سياق التحليلات (أيْ: سرد القصص القائم على البيانات). اقرأ كلًا من المقالة والورقة ثم قُمْ بتحليلهما بشكلٍ نقدي، واكتب تقريرًا يعكسُ فهمك وآراءك حول أهمية السرد القصصي في ذكاء الأعمال وتحليلات الأعمال.

١٢- انتقل إلى Data.gov - وهي بوابة بيانات ترعاها حكومة الولايات المتحدة - والتي تحتوي على عددٍ كبيرٍ جدًا من مجموعات البيانات حول مجموعة متنوعة من الموضوعات التي تتراوح من الرعاية الصحية إلى التعليم، والمناخ والسلامة العامة. اختر موضوعًا يشغل بالك كثيرًا. انتقل من خلال المعلومات الخاصة بهذا الموضوع والشرح المقدَّم على الموقع. استكشف إمكانيات تنزيل البيانات، واستخدم أداتك المفضلة للتصوير البياني للبيانات لإنشاء معلومات ومقاطع مرئية ذات معنًى خاص بك.

مهامُّ فريق ومشروعات لعب الأدوار:

١- يبدأ التحليل بالبيانات. إن تحديد البيانات ذات الصِّلة والوصول إليها والحصول عليها ومعالجتها، هي أهمُّ المهامِّ في أيِّ دراسة تحليلية. بصفتك فريقًا، يتمُّ تكليفك بالبحث عن بيانات واقعية كبيرة بما يكفي (إما من مؤسستك الخاصة، وهذا هو الخيار الأفضل، أو من الإنترنت والتي يمكن أن تبدأ ببحثٍ بسيطٍ، أو من روابط البيانات المنشورة على KDnuggets.com)، والتي لديها عشرات الآلاف من الصفوف وأكثر من ٢٠ متغيرًا لتتناول وتوثق المشروع الشامل للمعالجة المسبقة للبيانات. في معالجتك للبيانات؛ حَدِّد الشذوذات والتناقضات باستخدام طرق ومقاييس الإحصاء الوصفي، واجعل تحليلات البيانات جاهزةً. قم بإعداد قائمة مع التبرير لخطوات معالجتك المسبقة وقراراتك في تقرير شامل.

٢- انتقــل إلى موقع ويــب (dundas.com, idashboards.com, enterprise-dashboard.com) والمزوِّد بمعلومات لوحة المعلومات المعروفة جيدًا؛ إذ توفِّر هذه المواقع عددًا من الأمثلة على لوحــات التحكُّم التنفيذيــة. وبصفتك فريقاً، اختر صناعةً معينةً (مثــل: الرعاية الصحية، أو الخدمــات المصرفية، أو الخطوط الجوية). حَدِّد موقع بعــضٍ من الأمثلة للوحات معلومات هــذه الصناعــة. قم بوَصْف أنــواع المقاييس الموجودة في لوحات المعلومــات. ما هي أنواع شاشــات العرض التي يتمُّ اســتخدامها لتوفير المعلومات؟ وباســتخدام ما تعرفه عن تصميم لوحة المعلومات، قم بتوفير نموذج مبدئي لورقة لوحة معلومات لهذه المعلومات.

٣- انتقــل إلى teradatauniversitynetwork.com. ومن هناك؛ انتقل إلى مصادر بيانات جامعة أركنســاس. اختر واحدة من مجموعات البيانات الكبيرة، وقم بتنزيل عددٍ كبيرٍ من الســجلات (قــد يتطلَّب ذلك منك كتابة عبارة SQL التي تقوم بإنشــاء المتغــيرات التي تريد تضمينها في مجموعة البيانات). اطرح ١٠ أســئلة على الأقل يمكن معالجتها باستخدام تصوير المعلومات. ثم باستخدام أداتك المفضَّلة لتصوير البيانات (ولتكن Tableau)، حلل البيانات ثم قم بإعداد تقرير مُفصَّل يتضمن لقطات شاشة وصور مرئية أخرى.

المراجع:

- Abela, A. (2008). Advanced presentations by design: Creating communication that drives action. New York: Wiley.
- Annas, G. J. (2003). HIPAA regulations—A new era of medical-record privacy? New England Journal of Medicine, 348(15), 1486-1490.
- Ante, S. E., & McGregor, J. (2006). Giving the boss the big picture: A dashboard pulls up everything the CEO needs to run the show. Business Week, 43-51.
- Carotenuto, D. (2007). Business intelligence best practices for dashboard design. WebFOCUS white paper. www.datawarehouse.inf.br/papers/information_builders_dashboard_best_practices.pdf (accessed August 2016).
- Dell customer case study. Medical device company ensures product quality while saving hundreds of thousands of dollars. https://software.dell.com/documents/instrumentation-laboratory-medical-device-companyensures-product-quality-while-saving-hundreds-ofthousands-of-dollars-case-study-80048.pdf (accessed August 2016).
- Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. Decision Support Systems, 49(4), 498-506.
- Delen, D. (2011). Predicting student attrition with data mining methods. Journal of College Student Retention 13(1), 17-35.
- Delen, D., Cogdell, D., & Kasap, N. (2012). A comparative analysis of data mining methods in predicting NCAA bowl outcomes. International Journal of Forecasting, 28, 543-552.
- Delen, D. (2015). Real-world data mining: Applied business analytics and decision making. Upper Saddle River, NJ: Financial Times Press (A Pearson Company).
- Eckerson, W. (2006). Performance dashboards. New York: Wiley.
- Few, S. (2005, Winter). Dashboard design: Beyond meters, gauges, and traffic lights. Business Intelligence Journal, 10(1).
- Few, S. (2007). Data visualization: Past, present and future. perceptualedge.com/articles/Whitepapers/Data_Visualization.pdf (accessed July 2016).
- Fink, E., & Moore, S. J. (2012). Five best practices for telling great stories with data. White paper by Tableau Software, Inc., www.tableau.com/whitepapers/telling-data-stories (accessed May 2016).
- Freeman, K. M., & Brewer, R. M. (2016). The politics of American college football. Journal of Applied Business and Economics, 18(2), 97-101.
- Gartner Magic Quadrant, released on February 4, 2016, gartner.com (accessed August 2016).
- Grimes, S. (2009a, May 2). Seeing connections: Visualizations makes sense of data. Intelligent Enterprise.i.cmpnet.com/intelligententerprise/next-era-business-intelligence/Intelligent_Enterprise_Next_Era_BI_Visualization.pdf (accessed January 2010).

- Grimes, S. (2009b). Text analytics 2009: User perspectives on solutions and providers. Alta Plana. altaplana.com/TextAnalyticsPerspectives2009.pdf (accessed July, 2016).

- Hardin, M., Hom, D., Perez, R., & Williams, L. (2012). Which chart or graph is right for you? Tableau Software: Tell Impactful Stories with DataŒ. Tableau Software. http://www.tableau.com/sites/default/files/media/which_chart_v6_final_0.pdf (accessed August 2016).

- Hernández, M. A., & Stolfo, S. J. (1998, January). Real-world data is dirty: Data cleansing and the merge/purge prob- lem. Data Mining and Knowledge Discovery, 2(1), 9-37.

- Hill, G. (2016). A Guide to enterprise reporting. ghill.customer.netspace.net.au/reporting/definition.html (accessed July 2016).

- Kim, W., Choi, B. J., Hong, E. K., Kim, S. K., & Lee, D. (2003). A taxonomy of dirty data. Data Mining and Knowledge Discovery, 7(1), 81-99.

- Kock, N. F., McQueen, R. J., & Corner, J. L. (1997). The nature of data, information and knowledge exchanges in busi- ness processes: Implications for process improvement and organizational learning. The Learning Organization, 4(2), 70-80.

- Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised leaning. International Journal of Computer Science, 1(2), 111-117.

- Lai, E. (2009, October 8). BI visualization tool helps Dallas Cowboys sell more Tony Romo jerseys. ComputerWorld.

- Novell. (2009, April). Executive dashboards elements of success. Novell white paper. www.novell.com/docrep/documents/3rkw3etfc3/Executive Dashboards_Elements_of_Success_White_Paper_en.pdf (accessed June 2016).

- Quinn, C. (2016). Data-driven marketing at SiriusXM. Teradata Articles & News. at http://bigdata.teradata.com/US/Articles-News/Data-Driven-Marketing-At-SiriusXM/(accessed August 2016); Teradata customer success story. SiriusXM attracts and engages a new generation of radio consumers. http://assets.teradata.com/resourceCenter/downloads/CaseStudies/EB8597.pdf?processed=1.

- Radha, R. (2008). Eight best practices in dashboard design. Information Management. www.information-manage-ment.com/news/columns/-10001129-1.html (accessed July 2016).

- SAS. (2014). Data visualization techniques: From basics to Big Data. http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/data-visualization-techniques-106006.pdf (accessed July 2016).

- Thammasiri, D., Delen, D., Meesad, P., & Kasap N. (2014). A critical assessment of imbalanced class distribution prob- lem: The case of predicting freshmen student attrition. Expert Systems with Applications, 41(2), 321-330.

الفصل الثالث
التحليلات الوصفية (٢)
ذكاء الأعمال ومستودعات البيانات

أهداف التعلُّم:

- فهم التعريفات والمفاهيم الأساسية لمستودعات البيانات.
- فهم فنيات مستودعات البيانات.
- وَصْف العمليات المُستَخدَمة في تطوير وإدارة مستودعات البيانات.
- شَرْح عمليات مستودعات البيانات.
- شَرْح دور مستودعات البيانات في دَعْم القرار.
- شَرْح تكامل البيانات، وعمليات الاستخراج، والتحويل والتنزيل (ETL).
- فَهْم جوهر إدارة أداء الأعمال (BPM).
- تعلُّم بطاقة الأداء المتوازن وستة سيجما كنُظُم قياس الأداء.

لقد كان مفهومُ مستودعات البيانات موجودًا منذ أواخر الثمانينيات. يُقدِّم هذا الفصل الأساسَ لنوعٍ مهمٍّ من قواعد البيانات، يُسمَّى مستودع البيانات، والذي يُستَخدَم أساسًا لدعم القرار ويوفر الأساس المعلوماتي لتحسين القدرات التحليلية. وسنناقش مفاهيم تخزين البيانات وما يرتبط بذلك، وكذلك إدارة أداء الأعمال في الأقسام التالية.

٣-١ مقال افتتاحي: استهداف الاحتيال الضريبي مع ذكاء الأعمال، ومستودعات البيانات:

يتَعيَّن على الحكومات أن تعمل بجدٍّ لمنع التهرُّب الضريبي، والذي يؤدي إلى تناقص ملحوظ في عائداتها. ففي عام ٢٠١٣م؛ نجحت إدارة العائدات الداخلية (IRS) في إجهاض محاولات من مجهولين، كانت تستهدف خداع الحكومة الفيدرالية في مبلغ ٢,٢٤ مليار دولار من مبالغ الضرائب. ومع ذلك، وفي نفس العام، خسرت مصلحة الضرائب الأمريكية مبلغ ٥,٨ مليار دولار على مطالبات تمَّ إدراجُها في وقتٍ لاحقٍ تحت بند التهرُّب الضريبي.

كما تخسر الولايات الأمريكية أيضًا الكثير من الأموال عندما يستخدم المحتالون أرقام الضمان الاجتماعي المسروقة، ونماذج W-2، وغيرها من المعلومات الشخصية لتقديم مطالبات مزيفة. وقد زاد هذا النوع من الجرائم في السنوات الأخيرة بمعدلٍ يُنذر بالخطر. يقول بيتر فرانشوت، المراقب المالي بولاية ماريلاند: «لقد سمع كلُّ الأمريكيين تقريبًا عن سرقة الهوية؛ لكن القليل منهم فقط كان على درايةٍ بهذه الزيادة الهائلة في تزوير العائدات الضريبية. إن هذه المشكلة المزعجة تؤثر على كلِّ دول العالم. وهذا ما يُعرَف - حرفيًّا - بالسطو الممنهج على أموال دافعي الضرائب».

في ولاية ماريلاند؛ كان الأشخاص المتهمون بتزييف مطالبات الاسترداد أعضاءً في فريق الكشف عن العائد المشكوك فيه (QRDT). ومثل نظرائهم في العديد من الولايات الأخرى، يستخدم هؤلاء الخبراء البرمجيات؛ لتحديد العوائد المشبوهة. ثم يقومون بعد ذلك بالتحقيق في العائدات لتسليط الضوء بدقة على الاحتيالي منها.

التحَدِّي:

في الماضي؛ كانت ولاية ماريلاند تستخدم مقاييسَ لفحص الإقرارات الضريبية واحدًا تلو الآخر. فإذا اتَّسَم العائدُ بصفات مُحدَّدة - مثل: نسبة معينة من الأجور المكتسبة إلى الأجور المحتجزة - قامت البرمجيات بتعليقِ ذلك العائد لحين التحقيق فيه. ثم قام أعضاء QRDT بالنظر في كلِّ العائدات المعلقة؛ وذلك عن طريق مقارنة الأجر والمعلومات المحجوبة بالأرقام المقدَّمة من نموذج W-2 الذي تمَّ استلامه من قِبل صاحب العمل. كانت العملية كثيفة العمالة وغير فعالة. ومن أصل ما يقرب من ٢,٨ مليون عائد ضريبي كانت تحصل عليه ميريلاند كلَّ عام، قام أعضاء فريق QRDT بتعليق حوالي ١١٠٠٠٠ منها. ولكن تبَيَّن أن معظم تلك العائدات كانت عائدات مشروعة. يقول آندي شاوفيل، مدير مكتب تقديرات الإيرادات لمراقبة حسابات ولاية ماريلاند: «لقد تبَيَّن لنا أن ١٠٪ فقط من المموِّلين كانوا مخادعين».

في سنةٍ نموذجية؛ أنقذت تلك العملية ولاية ماريلاند من سداد ما بين ٥ إلى ١٠ ملايين دولار كمطالبات استرداد مزيفة. وعلى الرغم من نجاح ذلك؛ فإنها تُعَدُّ مجرد حلٍّ متواضع، مع الأخذ بعين الاعتبار الموارد المرتبطة بالعملية وإزعاج دافعي الضرائب المخلصين الذين تمَّ تحديد عوائدهم للتحقيق. يقول شاوفيل: «كان التفكير في أننا كنا نعلق ما بين ٩٠٠٠٠ و١٠٠٠٠٠ عملية استرداد للضرائب أمرًا صعبًا. لقد أردنا الحصول على هذه المبالغ المستردَّة لدافعي الضرائب بشكلٍ أسرع؛ إذ يعتمد الكثيرُ من الناس على هذا المال كجزءٍ من دخلهم».

الحل:

احتاجــت ماريلاند إلى عملية أكثر فعالية. كما احتاجت إلى إســتراتيجيات جديدة لاســتمرار السيطرة على المحتالين. يقول شاوفيل: «كانت جميع الولايات، وكذلك مصلحة الضرائب الداخلية IRS، يستخدمون جميعًا نفس المقاييس التي كنا نستخدمها، ولا أعتقد أنه كان من الصعب على المجرمين معرفة الإجراءات الاحترازية التي نتبعها». ولحُسْــن الحظ؛ حصلت ماريلاند مؤخرًا على ســلاحٍ جديدٍ قــويٍّ ضد التهرُّب الضريبي. في عام ٢٠١٠م، عمل مراقــب الخزينة في ماريلاند مع شركــة Teradata of Dayton، بولاية أوهايو، لتنفيذ مســتودع بيانــات مُصمَّم لدعم مجموعة متنوعة من مبادرات الامتثال.

وبينما كان المسؤولون يناقشون المبادرات التي يجب إطلاقها؛ طفت فكرةٌ واحدةٌ على السطح. تقــول شــارون بوناردي، نائبة مراقب ولايــة ماريلاند: «لقد قرَّرنا أننا يجــبُ أن نعطي الأولوية لجهودنا الرامية إلى تعقب الاحتيال في استرداد الأموال». لذا بدأت الولاية العمل مع تيراداتا ومع شركة ASR Analytics في بوتوماك بولاية ماريلاند؛ لتطوير عملية أفضل لعزل الإقرارات الضريبية الاحتيالية (TempleWest, 2013).

يقول شاوفيل: «كانت الخطوة الأولى؛ هي تحليل بياناتنا وتعلُّم ما يمكن تعلُّمه عن الاحتيال». ومن بين الاكتشافات الأخرى؛ أظهر التحليلُ أنه عندما يتم تعليق عائدات متعددة - حتى لأسباب مختلفة تمامًا - فإنها غالبًا ما تكون لها سمات مشتركة. قامت الولاية ببناء قاعدة بيانات للسمات التي تميِّز الخصائص الاحتيالية للعائدات، وكذلك للسمات التي يتميَّز بها الشرفاء. يقول شاوفيل: «لقــد عملنا مع ASR لوضع هذه المعلومات معًا وتطويــر الانحدارات الخطية، وبدلًا من النظر إلى المقاييــس المنفردة بذاتها؛ شرعنــا في تطبيق فكرة الجمع بين العديد من هذه المقاييس معًا». وكانت النتيجة صورةً أكثر دقة من حيث درجات الاختلاف للعائد الاحتيالي المعتاد.

وبــدلًا من تعاقب العائــدات الضعيفة واحدًا تلــو الآخر؛ يُحدِّد النظــام الجديد مجموعات العائدات التي تبدو مشــبوهةً لأسباب مماثلة. تجعل هذه الإستراتيجية عملية التحقيقات تسيرُ بشــكلٍ أسرع. ويُحــدِّد نظام التحليلات أيضًا درجــةً لكلِّ عائد؛ وذلك بنــاءً على مدى احتمالية كونه احتياليًّا. ثم ينتج بعد ذلك قائمة ذات أولوية لتوجيه سَــيْر عمل QRDT. يقول شــاوفيل: «نحن نعمل أولًا على العائدات التي من الأرجح ألا تكون احتيالية؛ ولذا يمكننا إخراجها من قائمة الانتظار». ثم تعود العائدات الأكثر شبهةً للمراجعة مرةً أخرى.

النتائج:

يقــول بوناردي: «بفضل نماذج التحليلات هذه؛ يمكننا تقليلُ القِيَم الموجبة الخاطئة، حتى لا نثقــل على كاهل دافعي الضرائب الذين قدَّموا إقراراتهم الضريبية بدقة إلى الولاية». وبمجرد أن يزيــلَ المحققون عائداتهم من قائمة الانتظار؛ يصبح بإمكان دافعي الضرائب اســترداد مطالباتهم التي حُصِّلَت منهم بالخطأ.

وبفضــل التقنية الجديدة؛ تتوقــع QRDT تعليق ما يتراوح من ٤٠٠٠٠ إلى ٥٠٠٠٠ فقط من العائدات الضريبية، مقارنةً بـ ١١٠٠٠٠ في الســنوات الماضية. يقول شاوفيل: «من بين كل هؤلاء الذين عملنا معهم حتى الآن؛ نحصل على نســبة دقة تصل إلى ٦٥٪ تقريبًا». ويُعَدُّ ذلك تحسُّــنًا رائعًا مقارنةً بمعدل النجاح التاريخي البالغ ١٠٪ فقط. ويقول فرانشوت، المراقب المالي لماريلاند: «بمجــرد أن يتــمَّ تحديدُ العائــدات التي قد تكــون احتياليةً، يمكن لفريق الخــبراء الممتحنين أن يراجعهــا بعناية، في وقتٍ واحد؛ لإزالة تلك العائدات التي تثبت عدم شرعيتها. وبذلك تتحسَّــن العملية برُمَّتها بشكلٍ أفضل طوال الوقت».

وفي أواخر مارس من نفس العام؛ ســاعدت التحليلات المتقدِّمة فريق QRDT على اســترداد حوالي ١٠ مليون دولار في موسم الإيداع الحالي. يقول شاوفيل: «في ظل النظام القديم، كان هذا الرقــم يبلــغ حوالي ٣ ملايين دولار في هذه المرحلة». ولا يقتــصر دور التقنية الجديدة فقط على تيسير عمل QRDT بشكلٍ أسرع وأكثر كفاءة؛ ولكنها أيضًا تساعد الفريق على التعامل مع عبء العمــل الأكثر تعقيدًا. وعندما زاد المتهربــون من الضرائب من جهودهم، لجأ فريق QRDT إلى نَشْر إستراتيجيات جديدة ضدهم. فعلى سبيل المثال، في عام ٢٠١٥م تلقَّى الفريق حوالي ١٠٠٠٠ إخطــار مــن دافعي الضرائب الذين تمَّت سرقة هوياتهم. يقول شــاوفيل: «إذًا لدينا سَــيْر عمل جديــد: وهو البحث عــن أرقام الضمان الاجتماعي الخاصة بهم ومحاولة العثور على أيِّ حوادث تزوير قد تكون أُرتكِبَت معهم، ويُعَدُّ هذا مستوًى جديدًا من الجهود التي تُكمِّلها هذه المجموعة الآن بدون موارد إضافية».

وللمحافظة على مُخططات ضريبية أكثر تطوراً، لا يقوم المحققون الآن بفحص استمارات W-2 الحالية فحســب؛ بل يقومون أيضًا بمقارنتها باستمارات نفس الممولين من السنوات السابقة؛ بحثًا عن التناقضات. ويضيف شاوفيل: «أصبحت التحقيقات أكثر تعقيدًا واستغراقًا لوقتٍ أطول، وإذا لم تتم مراجعة تلك الإجراءات بعناية؛ فستلاحقنا بلا شك بعض المشكلات الحقيقية».

أسئلة على المقالة الافتتاحية:

١- لماذا من المهمِّ لمصلحة الضرائب الأمريكية ولحكومات الولايات الأمريكية استخدام أدوات مستودعات البيانات، وأدوات ذكاء الأعمال في إدارة الدخل القومي؟

٢- ما هي التحَدِّيات التي تواجه ولاية ماريلاند فيما يتعلَّق بالتهرب الضريبي؟

٣- ما هو الحلُّ الذي اعتمدوه؟ وهل توافق على طريقتهم؟ ولماذا؟

٤- ما هي النتائج التي حصلوا عليها؟ وهل نجح الاستثمار في (ذكاء الأعمال) ومستودعات البيانات؟

٥- ما هي المشكلات والتحَدِّيات الأخرى التي تعتقد أن تواجهها الحكومات الفيدرالية وحكومات الولايات، والتي يمكن أن تستفيد من ذكاء الأعمال ومستودعات البيانات؟

ماذا يمكننا أن نتعلَّم من هذه المقالة الافتتاحية القصيرة؟

تُوضِّح المقالة الافتتاحية قيمة ذكاء الأعمال، وأنظمة دعم القرار، ومستودعات البيانات في إدارة الإيرادات الحكومية. مع تنفيذ مستودع البيانات الخاصة بهم، تمكَّنت ولاية ميريلاند من الاستفادة من أصول البيانات الخاصة بها؛ لاتخاذ قرارات أكثر دقة وفي الوقت المناسب حول تحديد الإقرارات الضريبية الاحتيالية. وبتنفيذهم لفكرة مستودع البيانات، تمكَّنَت ولاية ماريلاند من الاستفادة من أصول البيانات الخاصة بها بشأن اتخاذ قرارات أكثر دقة وفي الوقت المناسب؛ وذلك فيما يتعلق بتحديد الإقرارات الضريبية الاحتيالية. وقد أتاح دَمْج ومعالجة مجموعة واسعة من مصادر البيانات داخل مستودع بيانات موحد لولاية ماريلاند، تحديد إشارات/ قواعد/ سمات الاحتيال الضريبي من الحقائق التاريخية بصورة أو توماتيكية بدلًا من الاعتماد فقط على الطرق التقليدية؛ إذ تمَّ تطبيق قواعد التصفية القائمة على الحدس. وباستخدام مستودعات البيانات وذكاء الأعمال، تمكَّنَت ولاية ماريلاند من تقليل مُعدَّل القيم الموجبة الخطأ بشكلٍ ملحوظ (وبذلك تمَّ تخفيفُ العبء الضريبي على الممولين) كما تمكَّنَت من تحسين معدل دقة التنبؤ من ١٠٪ إلى ٦٥٪ (أي أكثر من ستة أضعاف التحسن في دقة تحديد الإقرارات الضريبية الاحتيالية). ويُعَدُّ الدرس الرئيس هنا هو أن مستودع البيانات المصمَّم والمنفَّذ بشكلٍ صحيح مع أدوات وتقنيات ذكاء الأعمال يمكن أن يؤدي إلى تحسُّن معنوي (سواء على مستوى الدقة أو التوقيت)؛ مما يؤدي إلى فوائد (مالية أو غير مالية) لأيِّ منظمة، بما في ذلك حكومات الولايات مثل ماريلاند.

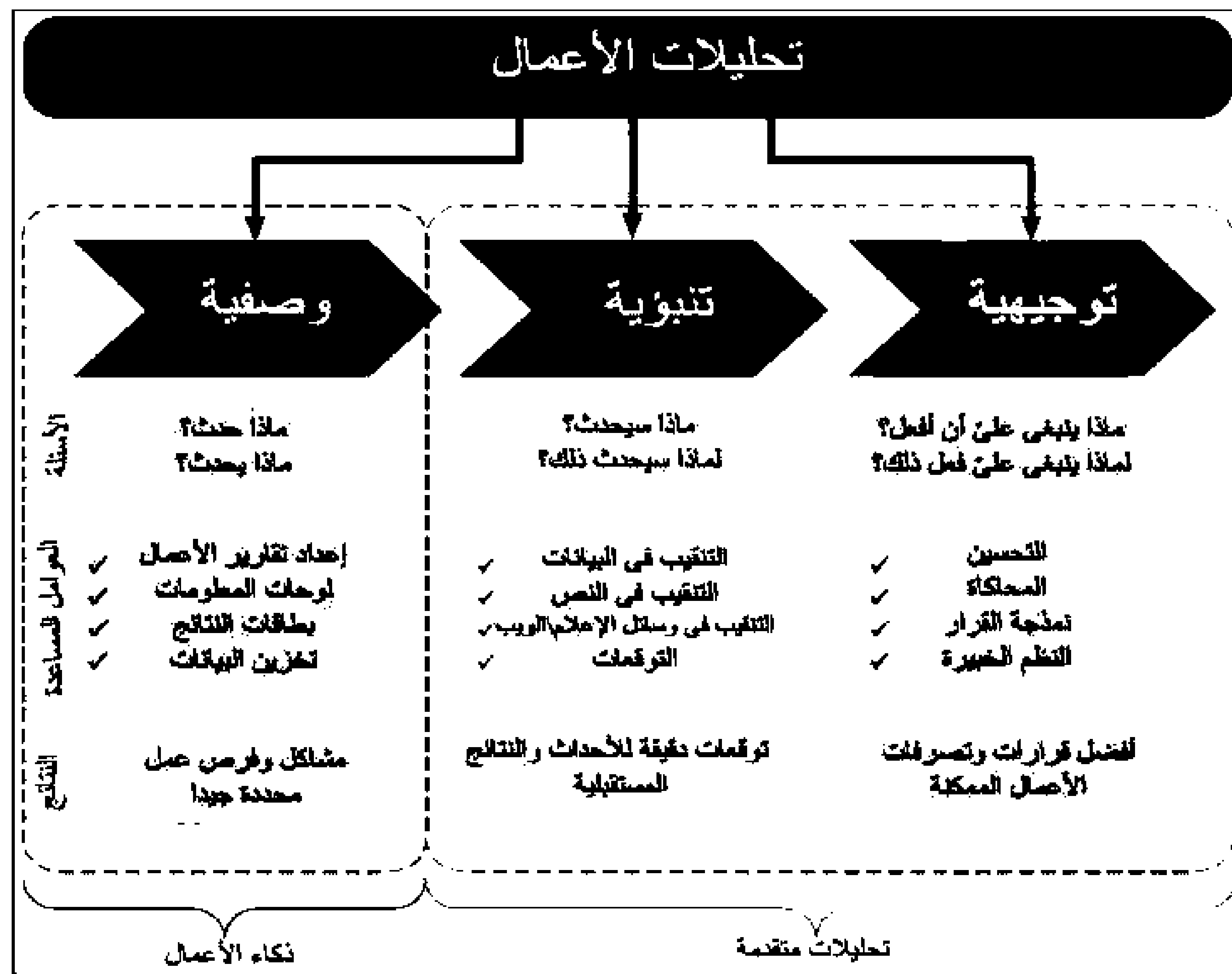
٣-٢ ذكاء الأعمال ومستودعات البيانات:

لقد كان ذكاءُ الأعمال، كمصطلح لوَصْف عملية اتخاذ القرار الإداري القائم على الحقائق/ الدليل، موجودًا منذ أكثر من ٢٠ عامًا. ومع ظهور تحليلات الأعمال ككلمة رنَّانة جديدة لوصف نفس الظاهرة الإدارية إلى حَدٍّ كبير؛ فقد انخفضت شعبية مصطلح ذكاء الأعمال. وعلى عكس كونه مصطلحاً شاملاً، يُستَخدَم ذكاء الأعمال في الوقت الحاضر لوصف المراحل المبكرة لتحليلات الأعمال (أيْ: التحليلات الوصفية).

يُوضِّح الشكل ٣-١ (والذي تمَّ عَرْض وشَرْح نسخةٍ مُبسَّطة منه في الفصل الأول لوصف تصنيف تحليلات الأعمال) العلاقة بين ذكاء الأعمال وتحليلات الأعمال من منظور مفاهيمي. وكما هو مُوضَّح هنا؛ فإن ذكاء الأعمال هو جزءٌ تحليليٌّ وصفيٌّ من سلسلة تحليلات الأعمال؛ إذ يؤدي نضجها إلى تحليلات متقدمة - وهي مزيجٌ من التحليلات التنبؤية والإلزامية.

تُعَدُّ التحليلات الوصفية (أيْ: ذكاء الأعمال) مستوًى مبدئيًّا للدخول في تصنيف تحليل الأعمال. وغالبًا ما يُطلَق عليها اسم إعداد تقارير النشاط التجاري؛ بسبب حقيقة أن معظم أنشطة التحليلات في هذا المستوى تتعامل مع إنشاء تقارير لتلخيص الأنشطة التجارية للإجابة عن أسئلة، مثل: «ماذا حدث؟» و«ما الذي يحدث؟». تشمل هذه الحزمة من التقارير لقطات ثابتة لمعاملات الأعمال التي تمَّ تسليمها إلى العاملين في مجال المعرفة (أيْ: صُناع القرار)؛ وذلك وفقًا لجدول زمني ثابت (يومي، أو أسبوعي، أو ربع سنوي)؛ وكذلك تقارير مُخصَّصة تُمكِّن صانع القرار من إنشاء تقريره الخاص (باستخدام واجهة مُستخدِم رسومية حدسية مُسطَّحة) لمعالجة حالة قرار مُحدَّدة أو فريدة من نوعها؛ وأيضًا وجهات نظر ديناميكية لمؤشرات أداء الأعمال الرئيسة (والتي غالبًا ما يتمُّ التقاطها وعرضها في نظام إدارة أداء الأعمال) يتمُّ تسليمها إلى المديرين والموظفين التنفيذيين في شكلٍ سهل الاستيعاب (مثل: واجهات رسومية تظهر كلوحة معلومات) بطريقة مستمرة.

Sources: Teradata case study. (2016). Targeting tax fraud with advanced analytics. http://assets.teradata.com/resourceCenter/downloads/CaseStudies/EB7183_GT16_CASE_STUDY_Teradata_V.PDF (accessed June 2016); Temple-West, P. (2013, November 7). Tax refund ID theft is growing "epidemic": U.S. IRS watchdog. Reuters. http://www.reuters.com/article/us-usa-tax-refund-idUSBRE9A61HB20131107 (accessed July 2016).



شكل ٣-١: العلاقة بين تحليلات الأعمال، وذكاء الأعمال، وتخزين البيانات

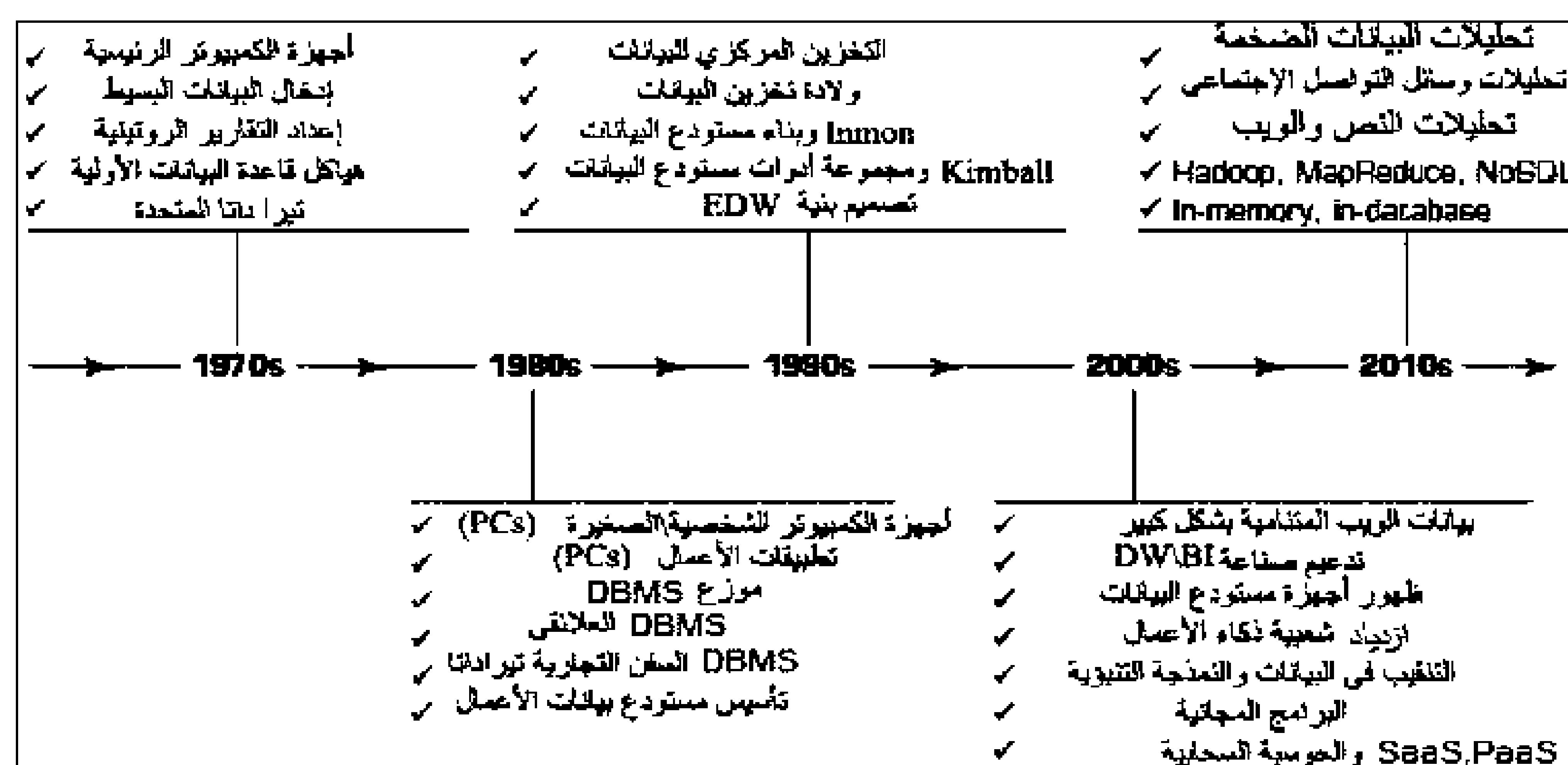
بشكلٍ عام؛ وكما هو مُبيَّن في الشكل (٣-١)، تعتمد أنظمة ذكاء الأعمال على مستودع البيانات كمصدرٍ للمعلومات؛ من أجل إنشاء ودَعْم القرارات الإدارية. إذ يتمُّ التقاط العديد من البيانات التنظيمية والخارجية وتحويلها وتخزينها في مستودع بيانات لدعم القرارات الدقيقة في الوقت المناسب من خلال نظرة عميقة للأعمال. ويهدف هذا الفصل إلى تغطية المفاهيم والطرق والأدوات المتعَلِّقة بمستودعات البيانات وإدارة أداء الأعمال.

ما هو مستودع البيانات؟

بعبارة بسيطة؛ يُعَدُّ مستودع البيانات (DW) تجمعاً من البيانات التي يتمُّ إنتاجها لدعم عملية صُنْع القرار. وهو أيضًا منطقةٌ غنيةٌ بالبيانات الحالية والتاريخية والتي من المحتمل أن تكون ذات أهمية كبيرة في المسقبل بالنسبة للمديرين في جميع أقسام المنظمة. وعادةً ما تكون البيانات مهيكلةً؛ بحيث تظهر في صيغة جاهزة لأنشطة المعالجة التحليلية (أي: المعالجة التحليلية عبر الإنترنت [OLAP]، التنقيب في البيانات، الاستعلام، إعداد التقارير، وغيرها من تطبيقات دعم القرار). إن مستودع البيانات؛ هو عبارة عن مجموعة موضوعية ومتكاملة وموجهة زمنيًّا وغير متطايرة من البيانات لدعم العملية الإدارية المتعلقة باتخاذ القرارات.

منظور تاريخي لمستودعات البيانات:

عــلى الرغم من أن مســتودعات البيانات هو مصطلح جديد نســبيًّا في تقنيــة المعلومات (IT)؛ فإنه من الممكن تتبُّع أصوله التاريخية، وحتى من قِبل التوسُّــع في اســتخدام أجهزة الحاســب. ففي أوائــل القرن العشرين؛ كان الناس يســتخدمون البيانات (بالطرق اليدويــة) لصياغة الاتجاهات التي تســاعد مســتخدمي الأعمال في اتخاذ قرارات مســتنيرة، وهو أكثر الأغراض الســائدة في مستودعات البيانــات. وترجــع - تاريخيًّا - الدوافع التي أدَّت إلى تطوير تقنيات مســتودعات البيانات إلى حقبة الســبعينيات، عندما هيمنت المركزية على عالم الحوســبة. وقد امتلكــت تطبيقات معالجة البيانات التجاريــة الحقيقية - والتي يتمُّ تشــغيلها على أجهزة الحاســب المركزية للــشركات - هياكل ملفات معقدة باســتخدام الجيل المبكر من قواعد بيانات (وليســت قواعد البيانات الجدولية المُســتَخدَمة في معظــم التطبيقــات في الوقــت الحالي) والتي يتمُّ إيــداع البيانات بها. وعــلى الرغم من أن هذه التطبيقــات قدَّمَت أداءً جيدًا في وظائف معالجة البيانات للمعاملات الروتينية؛ فإنَّ البيانات التي تمَّ إنتاجها من أداء هذه الوظائف (مثل: المعلومات المتعلقة بالعملاء، والمنتجات التي يطلبونها، والأموال التــي ينفقونها) قــد تمَّ دفنها بعيدًا في أعماق الملفات وقواعد البيانــات. وعندما ظهرت الحاجة إلى معلومات مُجمَّعة مثل: اتجاهات المبيعات بحسب المنطقة ونوع المنتج، كان ينبغي أن تطلب رسمياً من قســم معالجة البيانات؛ إذ تمَّ وَضْع ذلك الطلب في قائمة انتظار تضم مئات من طلبات التقارير الأخرى (Hammergren & Simon, 2009). وعلى الرغم من ظهور الحاجة إلى المعلومات والبيانات المُستَخدَمة لتوليد تلك المعلومات؛ فإنَّ تقنية قواعد البيانات لم تكن موجودةً لتحقيق ذلك. ويُوضِّح الشكل ٣-٢ جدولًا زمنيًّا يظهر فيه بعض الأحداث المهمة التي أدَّت إلى تطوير مستودعات البيانات.



شكل ٣-٢: قائمة الأحداث التي أدَّت إلى تطوير تخزين البيانات

في أواخــر القــرن الماضي؛ ظهرت شركات تجارية تعمل على تطويــر الأجهزة والبرمجيات، وقد شــهد ظهــور تلك الشركات بعض الحلول لهذه المشــكلة. وقد انبثق بــين عامي ١٩٧٦ و١٩٧٩م مفهــوم شركة جديدة - تيراداتا - من الأبحاث في معهد كاليفورنيا للتقنية (Caltech)، مســتندةً على مناقشــات أُجرِيَت مع مجموعة التقنية المتقدِّمة في Citibank. وقد عمل المؤسســون على تصميم نظــامٍ لإدارة قواعد البيانات للمعالَجة المتوازية مــع المعالجات الدقيقة المتعددة، والتي تستهدف بشكلٍ خاصٍّ دعم القرار. وفي ١٣ يوليو ١٩٧٩م، تمَّ تأسيسُ تيراداتا رسميًّا، وقد بدأت في جراج في Brentwood، بولاية كاليفورنيا. وقد تمَّ اختيار اســم تيراداتا لترمز إلى القدرة على إدارة تيرابايت (تريليونات من الحروف) من البيانات.

كانــت الثمانينيات هي عقد الحواسيب الشخصية والحواســيب الصغيرة. وقبــل أن يعرفها أحدٌ، لم تَعُد تطبيقات الحاســب الحقيقية موجودةً على أجهزة الحاسب المركزية فقط؛ بل كانت موجودةً في كلِّ مكانٍ يمكن أن تتواجد فيه أيُّ منظمة؛ مما تسبب في ظهور مُشكلةٍ خطيرة تُسمَّى جزر البيانات. وقد أدَّى حلُّ هذه المشكلة إلى نشأة نوعٍ جديد من البرمجيات، يُسمَّى نظام إدارة قاعدة البيانات الموزَّعة، والذي من شــأنه سَــحْب البيانات المطلوبة بطريقة ســحرية من قواعد البيانات من جميع أقســام المنظمة، وإعادة جميع تلــك البيانات مرةً أخرى إلى نفس المكان، ثم دمجها، وفرزها، وعمل كلِّ ما يلزم للإجابة عن ســؤال المســتخدِم. وعلى الرغم من كون المفهوم جيدًا وكذلك كون النتائج المبكرة من الأبحاث واعدة؛ فإنها كانت نتائج عادية وبســيطة؛ فهي لم تعمل بكفاءة على أرض الواقع، وبالتالي ظلت مشكلة جزر البيانات قائمة.

وفي الوقت نفسه؛ بدأت تيراداتا في شَحْذ كلِّ اهتماماتها تجاه المنتجات التجارية المتعلقة بحلِّ هذه المشكلة. وقد حصل بنك ويلز فارجو على أول نظام اختبار تيراداتا في عام ١٩٨٣م، وهو أحد أنظمــة إدارة قواعد البيانات العلائقية (RDBMS) وهو نظامٌ موازٍ لدعم القرار، ويُعَدُّ الأول من نوعه في العالم. وبحلول عام ١٩٨٤م، أصدرت تيراداتا نسخة من منتجها، وفي عام ١٩٨٦م، أطلقت مجلة Fortune على تيراداتا لقب «منتج العام». وقد قام منتج تيراداتا - والذي لا يزال قائمًا حتى اليوم - ببناء أول جهاز لمســتودعات البيانات، وهو عبــارة عن مجموعة من الأجهزة والبرمجيات لتلبية احتياجات مستودعات البيانات للكثير من المتعاملين في هذا الحقل. وبدأت كذلك العديدُ من الشركات الأخرى في صياغة إستراتيجياتها.

كــما وقعت عدة أحــداث أخرى خلال الثمانينيات من القرن الماضي؛ مما أكســبها لقب حقبة الابتكار في مســتودعات البيانات. فعلى ســبيل المثال: أسَّــس رالف كيمبال أنظمة Red Brick في عام ١٩٨٦م. بدأت أنظمة Red Brick في الظهور كشركة برمجيات رائدة من خلال مناقشة كيفية

تحسين الوصول إلى البيانات؛ ففي عام ١٩٨٨م؛ قدَّم كلٌّ من باري ديفلين وبول مورفي من IBM إيرلندا مصطلح مستودع بيانات الأعمال كمكوِّن أساسي لنظم المعلومات التجارية. وفي تسعينيات القرن العشرين؛ طفت على السطح طريقة جديدة لحل مشكلة جزر البيانات. وإذا كانت طريقة الثمانينيات قد فشلت في الحصول على البيانات والوصول إليها مباشرةً من الملفات وقواعد البيانات؛ فقد انطوت فلسفة التسعينيات على العودة إلى نهج السبعينيات؛ إذ تمَّ نَسْخُ البيانات من تلك الأماكن إلى مكانٍ آخر، ولكن يتمُّ فِعْل ذلك بشكلٍ أكثر دقة. ومن هنا وُلِدَت مستودعات البيانات. في عام ١٩٩٣م؛ كتب Bill Inmon كتاب بناء مستودع البيانات. يُصنِّف الكثيرُ من الناس Inmon كأب لمستودعات البيانات. وظهرت منشورات إضافية - بما في ذلك كتاب «مجموعة أدوات مستودع البيانات» لصاحبه رالف كيمبال ١٩٩٦م - والتي ناقشت تقنيات تصميم الأبعاد ذات الأغراض العامة لتحسين بنية البيانات لأنظمة دعم القرار المرتكزة على الاستعلام.

ومع بدايات الألفينات، وفي عالم مستودعات البيانات، استمرَّت شعبية البيانات في النمو. وبدأ اندماج مجتمع الموردين مع الخيارات. وفي عام ٢٠٠٦م؛ استحوذت شركة مايكروسوفت (Microsoft) على ProClarity؛ إذ قفزت إلى سوق مستودعات البيانات. وفي عام ٢٠٠٧م؛ قامت شركة أوراكل بشراء Hyperion، واستحوذت SAP على Business Objects، واندمجت IBM مع Cognos. لقد التهم قادة أكبر مُزوِّدي حلول نظم المعلومات في العالم قادة مستودعات البيانات في التسعينيات. وخلال هذه الفترة؛ ظهرت ابتكارات أخرى، بما في ذلك أجهزة مستودعات البيانات من مُورِّديها مثل Netezza (التي استحوذت عليها IBM)، وGreenPlum (التي دخلت في مظلة EMC)، وDATAllegro (التي اندمجت مع مايكروسوفت (Microsoft))، وأجهزة إدارة الأداء التي سهَّلت عملية مراقبة الأداء بشكلٍ فوري. وقد وفَّرت هذه الحلول المبتكرة وفورات في التكاليف؛ لأنها كانت متوافقة في تكوينها مع الحلول القديمة لمستودع البيانات.

منذ عام ٢٠١٠م؛ كانت الضجة الكبيرة هي البيانات الضخمة. يعتقد الكثيرون أن البيانات الضخمة ستؤثر في مستودعات البيانات. وحيال ذلك إما أنهم سيجدون طريقة للتعايش (والتي يبدو أنها الحالة الأكثر ترجيحًا، على الأقل لعدة سنوات) أو أن البيانات الضخمة (والتقنيات التي تصاحب وجودها) ستجعل التخزين التقليدي للبيانات أسلوبًا قديمًا ومُهمَلًا. وتتضمَّن التقنيات التي جاءت مع البيانات الضخمة كلًا من Hadoop وMapReduce وNoSQL وHive، وربما أننا سنكون بصدد ظهور مصطلح جديد تتمُّ صياغته في عالم البيانات؛ ليجمع بين ظهور مصطلحٍ جديد تمَّ صياغته في عالم البيانات الذي يجمع بين احتياجات وقدرات التخزين التقليدي للبيانات وظاهرة البيانات الضخمة.

خصائص مستودعات البيانات:

يُعَدُّ أحد الطرق الشـائعة لإدخال مسـتودعات البيانات؛ هو الرجوعُ إلى الخصائص الأساسية لتلك البيانات والتي يُعَدُّ من أهمِّها (انظر: Inmon, 2005):

- **التوجُّه بالموضوع:** يتمُّ تنظيمُ البيانات حسـب تفاصيل الموضوع، مثل: المبيعات، أو المنتجات، أو العملاء، والتي تحتوي فقط على المعلومات المتعلِّقة بدعم القرار. ولا يُمكِّن توجيه الموضوع المسـتخدمين من تحديد كيفية أداء أعمالهم فحسب؛ بل أيضًا يمكِّنهم من معرفة سبب أدائهم لتلك المهام. ويختلف مسـتودع البيانات عن قاعدة البيانات التشـغيلية؛ بحيث يكون لمعظم قواعد البيانات التشغيلية توجُّه بالمنتَج، ويتمُّ ضبطها للتعامل مع المعاملات التي تقوم بتحديث قاعدة البيانات. يُوفِّر اتجاه الموضوع نظرةً أكثر شمولية للمنظمة.
- **التكامل:** يرتبط التكامل ارتباطًا وثيقًا بتوجيه الموضوع. كما يجب على مسـتودعات البيانات أن توفـر البيانات القادمة من مصادر مختلفة في شـكلٍ متسـق. وللقيـام بذلك؛ فيجب على مسـتخدمي تلك المستودعات التعاملُ مع تسمية التعارضات والتناقضات بين وحدات القياس. ومن المفترض أن يكون مستودع البيانات متكاملًا تمامًا.
- **التغيُّر مع الوقت (سلسـلة زمنية):** يحتفظ المستودع بالبيانات التاريخية. وليس بالضرورة أن توفر البيانات الحالة الحالية (باستثناء الأنظمة الفورية). فهي تكتشف الاتجاهات والانحرافات والعلاقـات طويلة الأجل للتنبـؤ والمقارنات؛ مما يؤدي إلى اتخاذ قرار. ويكون لكلِّ مسـتودع بيانـات جودةٌ زمنية. ويُعَـدُّ الوقت هو البُعْد الأكثر أهميةً وربمـا الأوحد الذي يجب أن يتمَّ دعمـه من قِبل جميع مسـتودعات البيانات. كما تحتوي البيانـات المُجهَّزة للتحليل والقادمة من مصادر متعددة على نقاط زمنية مختلفة (كالمشاهدات اليومية، والأسبوعية، والشهرية).
- **عدم التطاير(الثبات):** فبُعْد مسـتودعات البيانات في المستودع، لا يُمكِّن للمستخدمين تغيير أو تحديث تلك البيانات؛ إذ يتمُّ تجاهل البيانات القديمة، ويتمُّ تسجيل التغييرات كبيانات جديدة.

هذه الخصائص السـالفة الذِّكر تمكِّن مستودعات البيانات ليتمَّ ضبطها بشكلٍ حَصْري تقريبًا لمهمة الوصول إلى البيانات. كما تُوجَد خصائص أخرى إضافية للبيانات والتي تتضمَّن ما يلي:

- **الارتكاز على شـبكة الإنترنت:** يتمُّ تصميمُ مستودعات البيانات عادةً لتوفير بيئة حوسبة فعَّالة للتطبيقات المستندة إلى الويب.
- **العلائقية/ متعددة الأبعاد:** يسـتخدم مسـتودعُ البيانات إمَّا بنيةً علائقيةً أو بنيةً متعددةَ الأبعاد. ويمكِن العثور على دراسةٍ استقصائية حديثة حول الهياكل المتعددة الأبعاد في روميرو وأبيلو (٢٠٠٩).

- **خدمة العملاء:** يَستخدمُ مستودع البيانات فنَّ خدمة العملاء؛ لتوفير وصول سهل للمستخدمين النهائيين.
- **الفورية:** توفِّر مســتودعات البيانات الأحدث إمكانيات الوصول إلى البيانات وتحليلها بشــكلٍ فوري أو نشط (انظر: Basu, 2003; Bonde & Kuckuk, 2004).
- **الاشــتمال على البيانات الوصفية:** يحتوي مســتودع البيانات على بيانات وصفية (بيانات عن البيانات) حول كيفية تنظيم البيانات وكيفية استخدامها بفعالية.

في الوقت الذي يُعَدُّ فيه مســتودعُ البيانــات مكاناً آمناً للبيانات، تُعَدُّ مســتودعات البيانات العمليــة برمتها حرفيًّا (انظــر: Watson، 2002). إذ تُعَدُّ مســتودعات البيانات نظامًا ينتج عنه تطبيقات توفر قدرةً على دعم القرار، كما يســمح بإمكانية الوصول إلى معلومات الأعمال، وأيضًا رؤيــة لتلك الأعمال. والأنواع الرئيســية الثلاثة لمخازن البيانات؛ هي: أســواق البيانات (DMs)، ومخازن البيانات التشغيلية (ODS)، ومستودعات بيانات المؤسسة (EDW). وفيما يلي مناقشةٌ لهذه الأنواع الثلاثة من المستودعات، ثم سنناقش أيضًا بعد ذلك البيانات الوصفية.

أسواق البيانات (DMs):

في الوقت الذي يضمُّ فيه مستودع البيانات قواعدَ البيانات الكبيرة عبر المؤسسة بأكملها، عادةً ما يكون ســوق البيانات (DM) أصغر حجمًا ويركِّز أكثر على موضوع أو قسم معين. إن DM هو عبــارة عن مجموعةٍ فرعيةٍ من مســتودع البيانات، تتكوَّن عادةً مــن منطقة موضوع واحد فقط (كالتســويق، أو العمليات). وقد تكون DM تابعةً أو مستقلةً. يكون سوق البيانات التابع عبارة عن مجموعة فرعية يتمُّ إنشاؤها مباشرةً من مستودع البيانات. كما أن لديها مزايا استخدام نموذج بيانات متســق وتوفير بيانات جيدة. وتقوم أســواق البيانات التابعة بدَعْم مفهوم نموذج بيانات واحد على مستوى المؤسسة بأكملها؛ ولكن يجبُ أولًا إنشاءُ مستودع البيانات. يضمن DM التابع للمســتخدِم النهائي أن يشاهد نفس الإصدار من البيانات التي يشاهدها كلُّ المستخدمين الآخرين لمســتودع البيانات. تتسبَّب التكلفة العالية لمستودعات البيانات في الحَدِّ من استخدامها للشركات الكبيرة. فتســتخدم العديدُ من الشركات بدلًا من ذلك إصدارًا منخفض التكلفة ومصغرًا لمستودع البيانات يُعرَف على أنه ســوق البيانات المســتقل. يُعَدُّ سوق البيانات المســتقل مستودعًا صغيرًا مُصمَّمًا لوحدة أعمال إستراتيجية؛ لكن لا يكون مستودع بيانات المشروع (EDW) مصدرًا له.

مخازن البيانات التشغيلية (ODS):

يُوفِّــر مخزن البيانات التشــغيلي (ODS) نموذجًا حديثًا إلى حَدٍّ مــا لملف معلومات العميل. وغالبًا ما يُستخدَم هذا النوع من قواعد البيانات كمنطقة مؤقتة لمستودعات البيانات. وبخلاف

المحتويــات الثابتة لمســتودع البيانــات، يتمُّ تحديــث محتويات (ODS) طوال فــترة العمليات التجارية. وتُستخدَم (ODS) في اتخاذ القرارات القصيرة الأجل التي تشمل التطبيقات ذات المهامِّ الحسَّاسة بدلًا من اتخاذ القرارات المتوسطة والطويلة الأجل المرتبطة بـ (EDW). تشبه (ODS) الذاكرة القصيرة المدى في أنها تخزِّن المعلومات الحديثة فقط. وفي المقابل؛ يشبه مستودع البيانات الذاكرة الطويلة المدى؛ لأنه يُخزِّن معلومات دائمة. يقوم ODS بدَمْج البيانات القادمة من أنظمة مصــادر مُتعدِّدة ويوفر رؤية متكاملة تقريبًا للبيانات المتغيرة والحالية. وتتطابق عمليات تبادل ونقل وتنزيل (ETL) - والتي ســتتم مناقشتها لاحقًا في هذا الفصل - الخاصة بـ (ODS) مع تلك العمليات المتعلقة بمســتودع البيانات. وأخيرًا، يتمُّ إنشاء سوق التشغيل (انظر: Imhoff, 2001) عند الحاجة إلى تحليل البيانات التشغيلية المتعددة الأبعاد. وتجدر الإشارة إلى أن مصدر البيانات الخاصة بسوق التشغيل هو نظام ODS.

مستودعات بيانات المؤسسة (EDW):

يُعَدُّ مســتودعُ بيانات المؤسسة (EDW) مستودعَ بيانات واسع النطاق يُستَخدَم لدعم القرار في جميع أنحاء المؤسســة. توفِّر طبيعة (EDW) الواســعة النطاق تكامــل البيانات القادمة من العديد من المصادر في شــكلٍ قياسيٍّ لتطبيقات ذكاء الأعمال الفعَّالة وتطبيقات دَعْم القرار. يتمُّ اســتخدامُ (EDWs) لتوفير بيانات للعديد من أنواع أنظمة دعم القرار (DSS)، بما في ذلك إدارة علاقــات العمــلاء (CRM)، وإدارة سلاســل التوريــدات (SCM)، وإدارة أداء الأعمال (BPM)، ومراقبة نشــاط العمــل، وإدارة دورة حياة المنتــج، وإدارة الإيرادات وأحيانًــا حتى أنظمة إدارة المعرفة. وتوضِّح الحالة العملية (٣-١) مجموعةً متنوعةً من المزايا التي تعود إيجابيًّا على شركات الاتصالات من جرَّاء تنفيذها لحلول تحليلات البيانات المخزَّنة في المستودع.

البيانات الوصفية:

تُعــرَّف البيانــات الوصفيــة بأنهــا بيانــات تصِــف بيانــات (عــلى ســبيل المثــال: انظــر Sen, 2004; Zhao, 2005). وتقــوم البيانات الوصفية بوَصْف بنية البيانات وبعض معانيها؛ مما يُســهِم في استخدامها الفعَّال أو غير الفعال. أشار ميهرا (٢٠٠٥م) إلى أن عدداً قليلاً من المنظمات تتفهَّــم بالفعــل طبيعة البيانــات الوصفية، ويتفهم عــددٌ أقل من الأول لكيفيــة تصميم وتنفيذ إســتراتيجية للبيانات الوصفية. ويتمُّ تعريفُ البيانات الوصفية بشــكلٍ عام؛ من حيث الاستخدام كبيانــات وصفية تقنيــة أو بيانات أعمال. وتُعَدُّ الأنماط طريقةً أخرى لعرض البيانات الوصفية؛ إذ

يُمكِننـــا من منظور نمطي أن نُفرِّق بين كلٍّ من البيانات الوصفية النحوية (أيْ: البيانات التي تَصِف بناء جملة البيانات) والبيانات الوصفية الهيكلية (أيْ: البيانات التي تَصِف بنية البيانات) والبيانات الوصفية ذات الدلالات اللفظية (أيْ: البيانات التي تَصِف معنى البيانات في مجال معين).

حالة عملية ٣-١

خطة بيانات أفضل: تســتفيد شركات الاتصالات الجيدة التأســيس من مســتودعات البيانات والتحليلات؛ للبقاء على القمة في صناعة تنافسية

جَنَــت شركاتُ الاتصــالات - التي ســاعدت على تحفيز النمو الهائــل لهذه الصناعة في أواســط وأواخر تســعينات القرن الماضي- منذ فترةٍ طويلة فوائدَ أن تكون أول ســوق (أيْ: الريــادة في هــذا الحقل). ولكن للحِفاظ على القدرة التنافســية؛ يجب على هذه الشركات تحســين كلِّ شيء باستمرار بدءًا من خدمة العملاء إلى تخطيط الأسعار. وفي الواقع؛ تواجه شركات الاتصــالات الجوالــة المخضرمة العديدَ من نفس التحَدِّيــات التي تواجهها الشركات الصاعدة، مثل: الاحتفاظ بالعملاء، وخفض التكاليف، وتحسين نماذج التسعير، ورَفْع مستوى رضا العملاء، واكتساب عملاء جُدد، وفَهْم دور وسائل الإعلام الاجتماعية في ولاء العملاء.

تلعبُ تحليلاتُ البيانات العالية الاســتهداف دورًا أكثر أهميةً في مســاعدة شركات الاتصالات الجوالة على تأمين أو تحسين مكانتها في سوقٍ تتزايد فيه المنافسة. وفيما يلي: كيف يعملُ بعض كبار مُزوِّدي الخدمات في العالم على إنشــاء مســتقبل قوي قائم على الأعمال التجارية القوية وذكاء العملاء.

المحافظة على العملاء:

قــد لا يخفى على أحد أن السرعة والنجاح اللذيــن يتعامل بهما مُقدِّم الخدمة مع طلبات الخدمة يؤثران بشكلٍ مباشرٍ على مستوى رضا العملاء، وبالتالي الميل إلى التسرب. لكن التحَدِّي هنا هو الوصول إلى العوامل التي لها أكبر الأثر.

تــدوِّن روكســان جارســيا - مديــرة مركــز العمليــات التجاريــة لشركــة Telefónica de Argentina - في ملاحظاتها: «إذا تمكَّنَّا من تتبُّع الخطوات المتبعة مع كلِّ عملية؛ فيمكننا أن نفهم نقاط الفشــل والتســارع، كما يمكننا قياسُ تدفقات العمل داخل الوظائف وفيما بينها، والتنبؤ بمؤشرات الأداء بدلًا من الاستجابة لها، ورفع مستوى الرضا العام عن تشغيل عملاء جدد».

كان حـلُّ الشركـة؛ هو مـشروع التتبُّع الذي بـدأ بـ ١٠ لوحـات معلومات في عام ٢٠٠٩م. ومنـذ ذلـك الحين؛ حقَّقَت الشركة عائدات سـنوية قدرهـا ٢,٤ مليون دولار أمريكي ووفورات في التكاليف، وتقليص وقت اكتساب العملاء الجدد، وانخفاض في عدد تسرب العملاء بنسبة ٣٠٪.

تقليل التكاليف:

يعتمد البقاءُ على القمة في أيِّ صناعة، في جزءٍ كبيرٍ منه، على الحفاظ على التكاليف في أدنى مسـتوياتها. بالنسبة لشركة Bouygues Telecom الفرنسية، تمَّ خَفْضُ التكلفة بصـورة آلية. يقـوم Aladin - وهو نظـامُ إدارة عمليات التسـويق القائم على شركة تيراداتـا - بنقـل عمليات الإنتاج التسـويقية المباشرة إلى الصورة الآليـة. وقد تخَطَّت مدخراتهـا في عامٍ واحدٍ أكثر من مليون دولار؛ في حـين تضاعفت تكاليف حملة البريد الإلكتروني الخاصة بها وإنتاج المحتوى ثلاثة أضعاف.

تقول كاثرين كـورادو - مديرة المشروع الرائدة في مجال المبيعات ومديرة اتصالات التجزئـة: «إنَّ هدفنا هو أن تكون أكثر إنتاجيةً واسـتجابةً لتبسـيط العمل الجماعي، ولتوحيد وحماية خبراتنـا، يجعل نظام Aladin أعضاء الفريق يركِّزون أكثر على القيمة المضافة للعمل؛ مـن خلال تقليل المهامِّ ذات القيمة المنخفضة؛ فتكون النتيجة النهائية المزيد من الجودة والإبداع في المخرجات».

هناك فائدة غير مقصودة من Aladin؛ ولكنها مُرحَّبٌ بها للغاية، وهي أنه تمَّ إلهامُ الإدارات الأخرى ببَدْء نَشْر مشاريع مماثلة لكلِّ شيء بدءًا من دعم مركز الاتصال وانتهاءً بعمليات إطلاق المنتجات/ العروض الترويجية.

اكتساب العملاء:

مع اختراق السـوق بنسـبة ١٠٠٪ أو أكثر في العديد من الدول، وبفضل المستهلكِين الذين يمتلكون أجهزةً متعددةً؛ فإنَّ مسألة كَسْب عملاء جُدُد لا تمثِّل تحدياً صغيراً. كما تواجه شركة Mobilink - أكبر شركة نقل في باكستان - صعوبة التشغيل في سوقٍ يمتلك فيه ٩٨٪ من المسـتخدمين خطةً مسـبقة الدفع، والتي تتطلَّب مشـتريات منتظمة من الدقائق الإضافية.

يقـول عمر أفضال - كبير المديرين في بي آي: «إنَّ رفع القيمة، على وجه الخصوص، يحافـظ على قوة وحساسـية دور الإيرادات في نموٍّ شركتنا؛ ففي السـابق كنَّا نفتقرُ إلى

القـــدرة على تعزيز هذا الجانــب من النموِّ التدريجي. لقد منحنــا نموذجنا لمعلومات المبيعات تلك القدرة؛ لأنها ســاعدت فريق المبيعات على رَسْــم خطط المبيعات القائمة على الإستراتيجيات الأكثر ذكاءً والمستندة على البيانات والتي تساعدنا على تزويد مُورِّدينا ببطاقات الـ SIM وكروت الخدش وقدرات زيادة الرَّصيد الإلكتروني بشكلٍ كامل».

ونتيجةً لذلك؛ لم تكسب Mobilink زيادةً في عدد المشتركين بنسبة ٢٪ فحسب؛ بل زادت أيضًا من اكتساب عملاء جدد بنسبة ٤٪ وتحسين ربحية تلك المبيعات بنسبة ٤٪.

شبكات التواصل الاجتماعي:

يعمل التوسُّــع في استخدام شــبكات التواصل الاجتماعي على تغيير عدد المنظمات التــي تتعامل مع كلِّ شيء بدءًا من خدمة العملاء، مرورًا بالمبيعات وانتهاءً بالتســويق. ويتجــه المزيدُ من شركات الاتصالات إلى تحويل اهتمامها بالشــبكات الاجتماعية؛ لفهم سلوك العملاء والتأثير فيه بشكلٍ أفضل.

وقد بادرت Mobilink بطَرْح مشروعٍ لتحليل الشــبكات الاجتماعية، والذي سوف يُمكِّن الشركة من استكشاف مفهوم التسويق الفيروسي وتحديد المؤثرين الرئيسين الذين يُمكِنهــم العملُ كســفراء علامة تجارية لبيع المنتجــات. وتبحث Velcom عن مؤثرين رئيســين مشــابهين إضافةً إلى عملاء منخفضي القيمة؛ إذ يُمكِن الاســتفادة من قِيمتهم الاجتماعية لتحســين العلاقات القائمة. وفي نفس الوقــت؛ تتطلَّع Swisscom إلى دَمْج جانب الشبكة الاجتماعية لسلوك العملاء مع بقية تحليلها خلال الأشهر القليلة القادمة.

الارتفاع إلى مستوى التحَدِّي:

عــلى الرغم من أن كلَّ ســوق يُقدِّم تحَدِّيات فريدةً خاصــةً به؛ فإنَّ معظم شركات الاتصــالات الجوَّالة تنفق الكثير من الوقت والموارد في إنشــاء ونَشْر وتنســيق الخطط لمواجهة كلٍّ من التحَدِّيات المُوضَّحة هنا. والخبر السَّار هو أنه مع اتساع وتطوُّر الصناعة وتقنية الهاتف المحمول على مَرِّ الســنين؛ فإنَّ حلول تحليلات البيانات التي تمَّ إنشاؤها لمواجهة هذه التحَدِّيات تُعَدُّ قادرةً على تلك المواجهة.

يقــوم تحليلُ البيانات الصوتية باســتخدام العمــلاء الحاليين، والأعــمال التجارية، ومعلومات الســوق للتنبؤ والتأثير على السلوكيات والنتائج المستقبلية. وتكون النتيجة النهائية؛ هي الوقوف على نَهْج أكثر ذكاءً، وأكثر رشاقةً، وأكثر نجاحًا للحصول على حصة سوقية وتحسين الرِّبحية.

أسئلة للمناقشة:

١- ما هي التحَدِّيات الرئيسية لشركات الاتصالات؟

٢- كيف يُمكِن لمســتودعات البيانات وتحليلات البيانات مساعدة شركات الاتصالات في التغلُّب على تحَدِّياتها؟

٣- لماذا تعتقد أن شركات الاتصالات مناسبة تمامًا للاستفادة الكاملة من تحليلات البيانات؟

Source: Marble, C. (2013). A better data plan: Well-established TELCOs leverage analytics to stay on top in a competitive industry. Teradata Magazine. http://www.teradatamagazine.com/v13n01/Features/A-Better-Data-Plan (accessed June 2016).

أسئلة مراجعة على القسم ٣-٢:

١- ما هو مستودع البيانات؟

٢- كيف يختلف مستودع البيانات عن قاعدة بيانات المعاملات؟

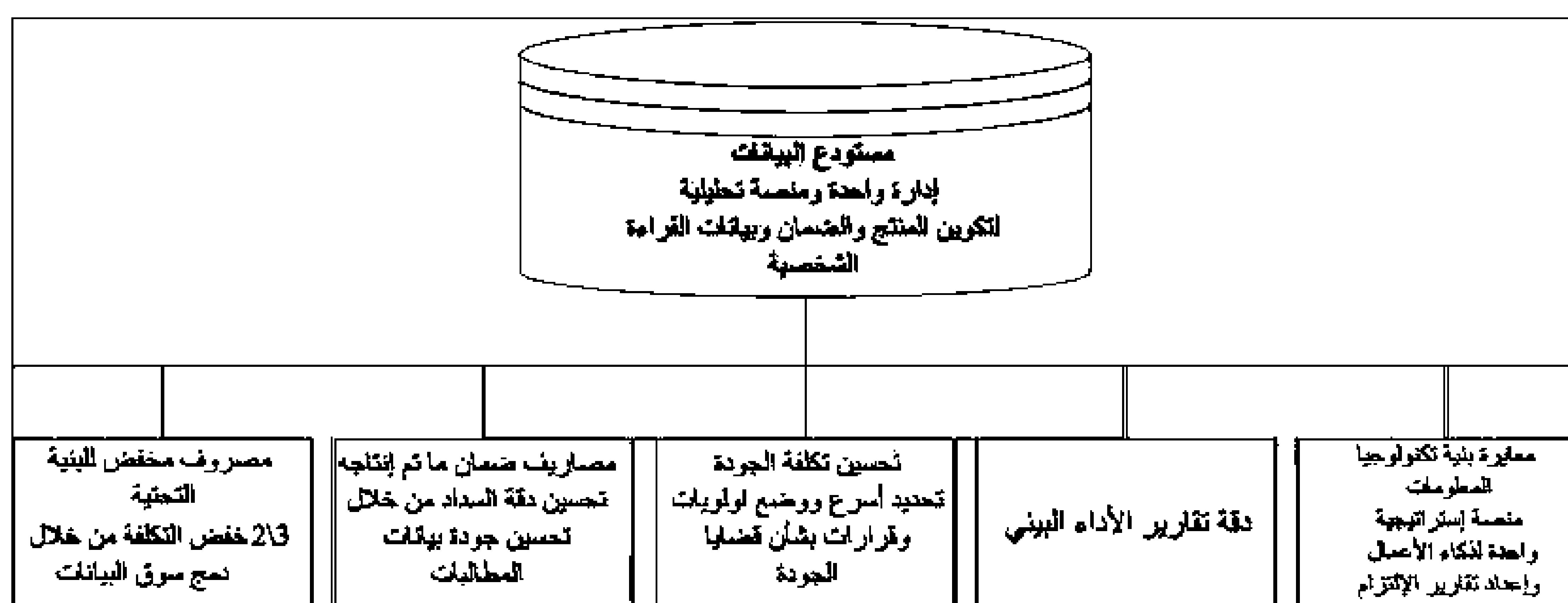
٣- ما هو ODS؟

٤- اذكر أوجه الاختلاف بين كلٍّ من DM، وODS، وEDW.

٦- ما هي البيانات الوصفية؟ اشرح أهمية البيانات الوصفية.

٣-٣ عملية مستودعات البيانات:

تقوم المؤسســاتُ - العامة والخاصة - بجَمْع البيانات والمعلومات والمعرفة باستمرار بمعدل متزايد السرعة وتخزينها في أنظمة محوسبة. وهنا يصبح الحِفاظ على هذه البيانات والمعلومات واستخدامها أمرًا معقدًا للغاية؛ خاصةً مع ظهور مشكلات التوسُّع. إضافةً إلى ذلك؛ يستمرُّ تزايد عدد المســتخدِمين الذين يحتاجون إلى الوصول إلى المعلومات نتيجةً لارتفاع مستوى المصداقية وســهولة الوصول إلى الشــبكة، وخاصة الإنترنت. لقد أصبح العمل مع قواعد بيانات متعددة - ســواء ذات مســتودع بيانات متكامل أم لا - مهمة بالغة الصعوبة تتطلَّب خبرةً كبيرةً؛ ولكنها يمكِن أن توفر فوائد هائلة تتجاوز تكلفتها بكثير. وكمثال توضيحي، يعرض الشــكل ٣-٣ فوائد الأعمال التجارية لمســتودع بيانات المؤسســة (EDW) التي قامت تيراداتــا ببنائه لصالح شركة تصنيع سيارات رائدة.



Source: Tradata Corp

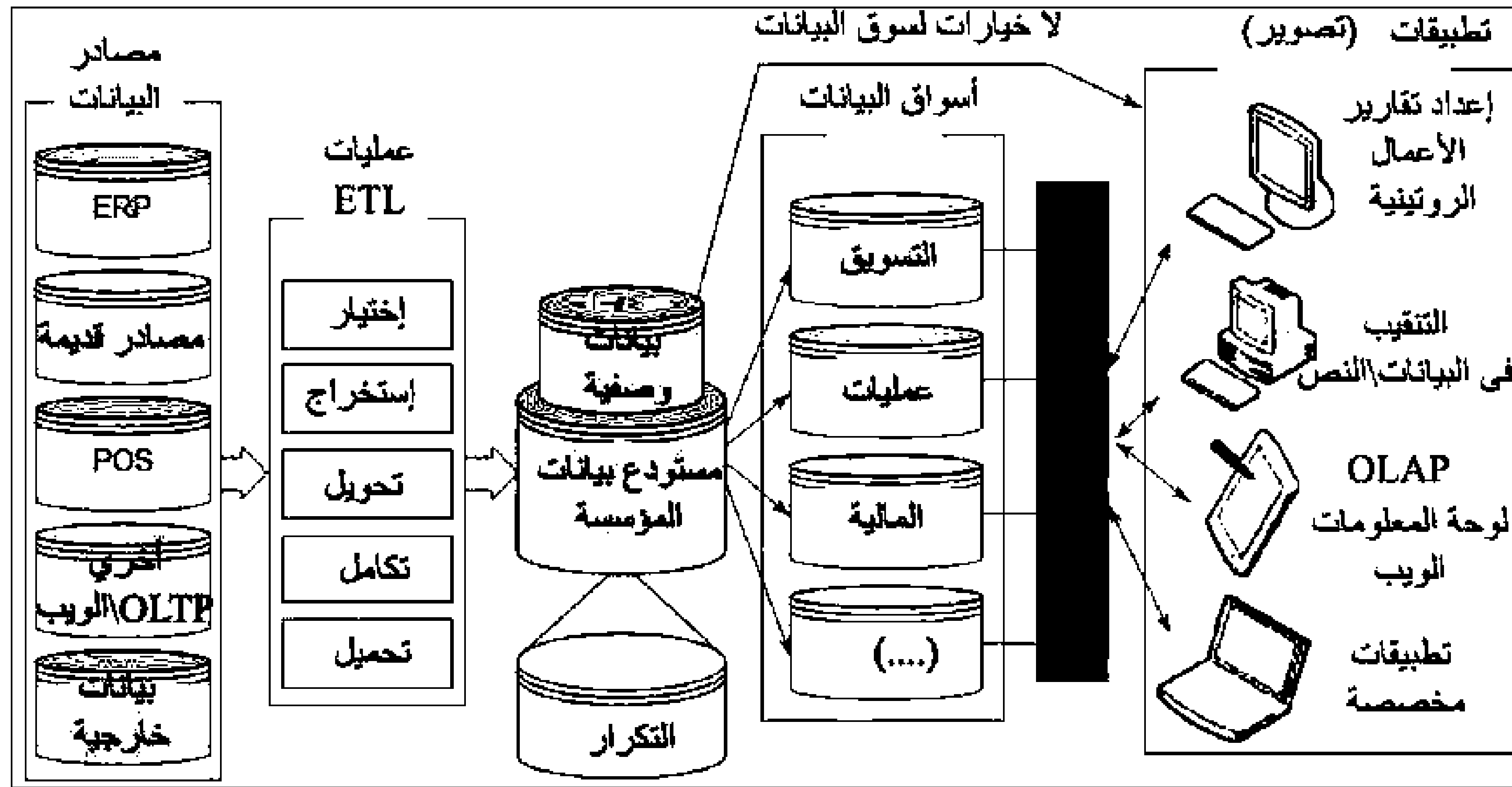
شكل ٣-٣: اتخاذ القرارات المعتمدة على البيانات - فوائد الأعمال لمستودع البيانات

تحتــاجُ العديدُ من المنظمات إلى إنشــاء مســتودعات البيانات، أيْ: مخــازن بيانات ضخمة لبيانات السَّلاسل الزمنية المتعلِّقة بدَعْم القرار. يتمُّ استيرادُ البيانات من مختلف الموارد الخارجية والداخلية، ويتمُّ تطهيرها وتنظيمها بطريقةٍ تتوافق مع احتياجات المؤسسة. وبعد مَلء البيانات في مســتودع البيانات؛ يمكن تنزيل أســواق البيانات (DMs) لمنطقةٍ ما أو قسم معين. وبدلًا من ذلك؛ يمكن إنشــاء DM أولًا - وفقًا للحاجة - ومن ثم دَمْجُها في EDW. ومع ذلك؛ فغالبًا لا يتمُّ تطوير DMs، وإنما يتمُّ تنزيل البيانات ببساطة على أجهزة الحاسب الشخصية أو تركها في حالتها الأصلية؛ من أجل معالجة بارعة مباشرةً باستخدام أدوات ذكاء الأعمال.

نعرض في الشكل ٣-٤ مفهوم مستودع البيانات. وفيما يلي المكوِّنات الرئيسية لعملية مستودعات البيانات:

- **مصادر البيانات:** يتمُّ الحصول على البيانات من عِدَّة أنظمة تشــغيلية مســتقلة «قديمة» وربما من مُزوِّدي البيانات الخارجيين (مثل: التعداد الســكاني الأمريكي). وقد تأتي البيانات أيضًا من (OLTP) أو نظام تخطيط موارد المؤسسات (ERP). وقد تتغَذَّى أيضًا بيانات الويب في شكلِ سجلات ويب من مستودع البيانات.
- **التنقيــب في البيانات وتحويلها:** يتمُّ التنقيبُ في البيانات وتحويلها بشــكلٍ صحيحٍ باســتخدام برنامج مكتوب خصيصًا أو مُعَدٍّ تجاريًّا لهذا الغرض ويُسمَّى ETL.
- **تنزيــلُ البيانــات:** يتمُّ تنزيلُ البيانــات في منطقة تدريجية؛ إذ يتــمُّ تحويلها وتطهيرها. تكون البيانات بعد ذلك جاهزةً للتنزيل في مستودع البيانات أو في DMs.

- **قاعدة بيانات شـــاملة:** وتُعَدُّ بشـــكلٍ أســـاسيٍّ هي (EDW) لدعم تحليل القرار بصورة شاملة؛ من خلال توفير معلومات ذات صِلة موجزة ومُفصَّلة قادمة من العديد من المصادر المختلفة.
- **البيانـــات الوصفية:** يتـــمُّ الاحتفاظ بالبيانات الوصفية بالصورة التي تســـمحُ بتقييمها من قِبل موظفي ومُستخدِمي التقنية. وتتضمَّن البيانات الوصفية برامج عن البيانات والقواعد المتعلقة بتنظيم ملخصات البيانات التي يَسهُل فهرستها والبحث فيها؛ خاصةً مع أدوات الويب.
- **الأدوات الوسيطة:** تتيحُ الأدوات الوسيطة الوصول إلى مستودع البيانات. ويمكِن للمستخدِمين المحترفين - كالمحللين مثلًا - كتابة اســـتعلامات SQL الخاصة بهم. يمكن للآخرين توظيف بيئة اســـتعلام يتمُّ إدارتها بصورةٍ جيـــدة، مثل كائنات الأعمال؛ وذلك بغـــرض الوصول إلى البيانات. هنـــاك العديدُ من التطبيقات ذات الواجهات الأمامية التي يُمكِن لأصحاب الأعمال التفاعل من خلالها مع البيانات المخزَّنة في مســـتودعات البيانات، بما في ذلك التنقيب في البيانات، OLAP، أدوات إعداد التقارير، وأدوات عرض البيانات.



شكل ٣-٤: إطار عمل وطرق عرض مستودع بيانات

أسئلة مراجعة على القسم ٣-٣:

١- صِف عملية مستودعات البيانات.

٢- صِف المكوِّنات الرئيسية لمستودع البيانات.

٣- حَدِّد وناقش دور الأدوات الوسيطة.

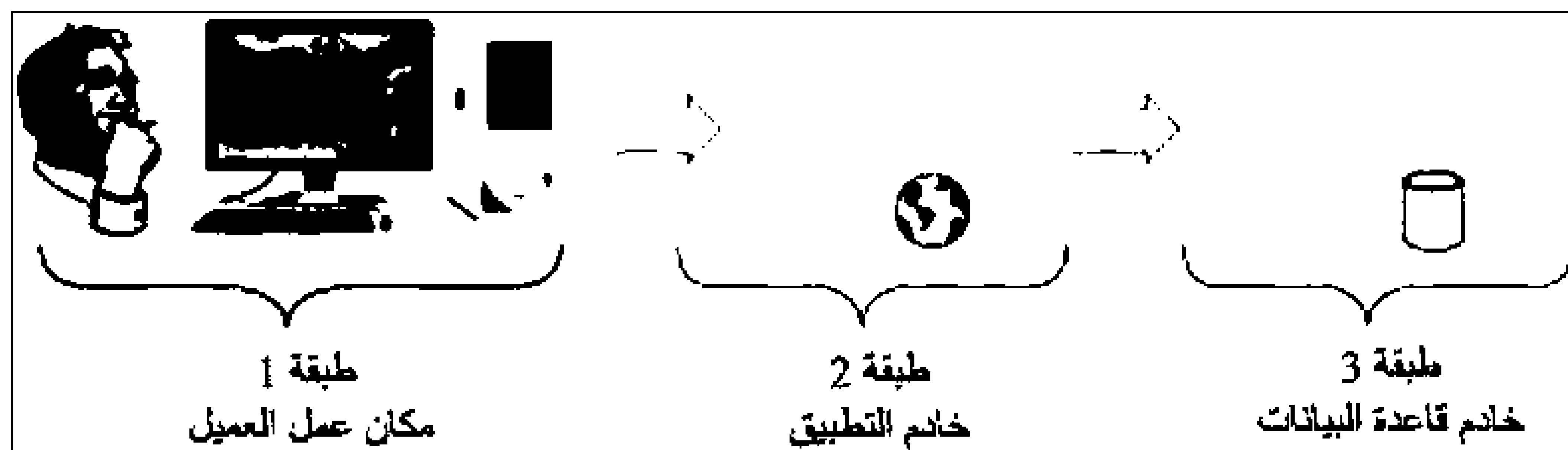
٣-٤ بنيات مستودعات البيانات:

يُمكِن اســتخدامُ العديدِ من فنيات نظام المعلومات الأساسية لمستودعات البيانات. وبشكلٍ عام؛ عادةً ما تُســمَّى هذه الفنيات العميل/ الخادم أو البنية ذات عدد (ن) الطبقات، والتي تكون أكثرها شيوعًا ثنائية أو ثلاثية الطبقات (انظر الشكلين ٣-٥ و٣-٦)؛ ولكن في بعض الأحيان يكون هناك طبقةٌ واحدةٌ فقط. ومن المعروف أنَّ هذه الأنواع من البنيات المتعددة الطبقات قادرةٌ على تلبية احتياجات أنظمة المعلومات الواســعة النطاق، والتي تتطلب الأداء، مثل: مســتودعات البيانات. وبالإشــارة إلى استخدام الفنيات المتعددة الطبقات المتداولة في مجال مستودعات البيانات؛ فإن هوفر، وبريسكوت، ومكفادين (٢٠٠٧) فرَّقوا بين هذه الفنيات من خلال تقسيم مستودع البيانات إلى ثلاثة أجزاء:

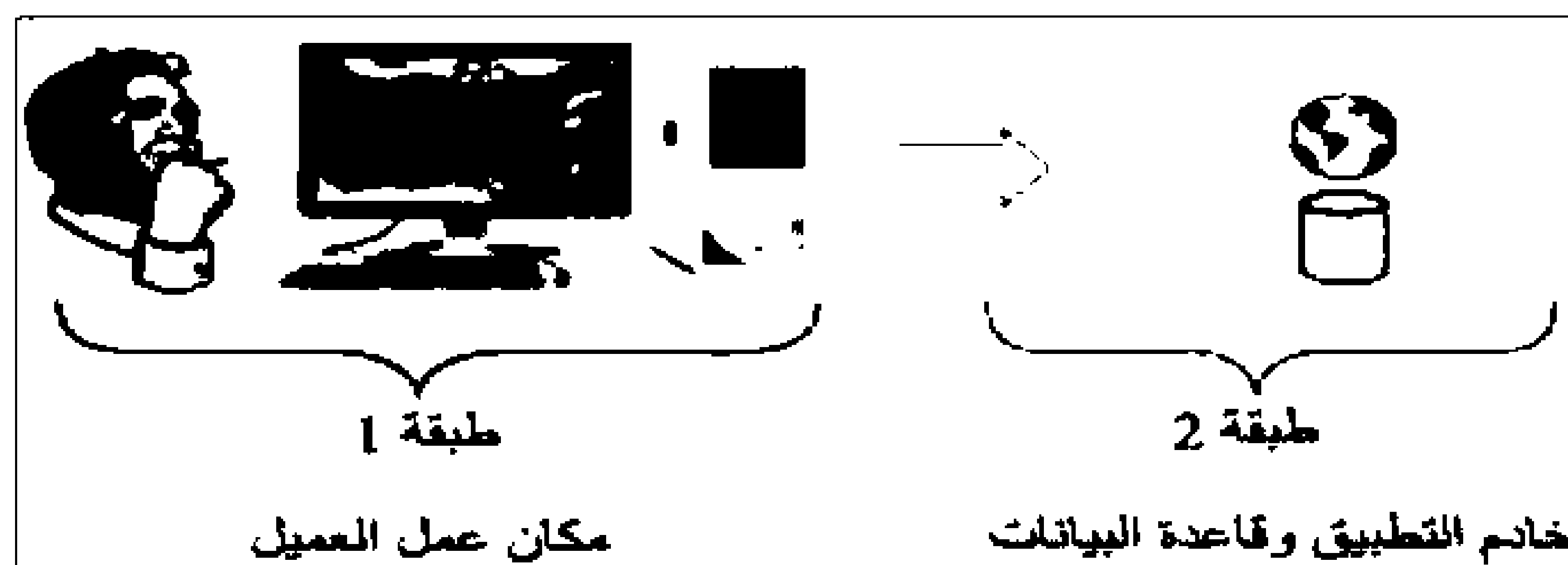
١- **مستودع البيانات نفسه:** والذي يحتوي على البيانات والبرامج المرتبطة به.

٢- **برنامج الحصول على البيانات (النهاية الخلفية):** والذي يستخرج البيانات من الأنظمة القديمة والمصادر الخارجية، ومن ثمَّ يقوم بدَمْجها وتلخيصها وتنزيلها في مستودع البيانات.

٣- **برنامج العميل (الواجهة الأمامية):** والذي يســمح لجميع المستخدمين بالوصول إلى البيانات وتحليلها من المستودع (a DSS/bi/business analytics [BA] engine).



شكل ٣-٥: بنية مستودع البيانات ثلاثية الطبقات



شكل ٣-٦: بنية مستودع البيانات ثنائية الطبقات

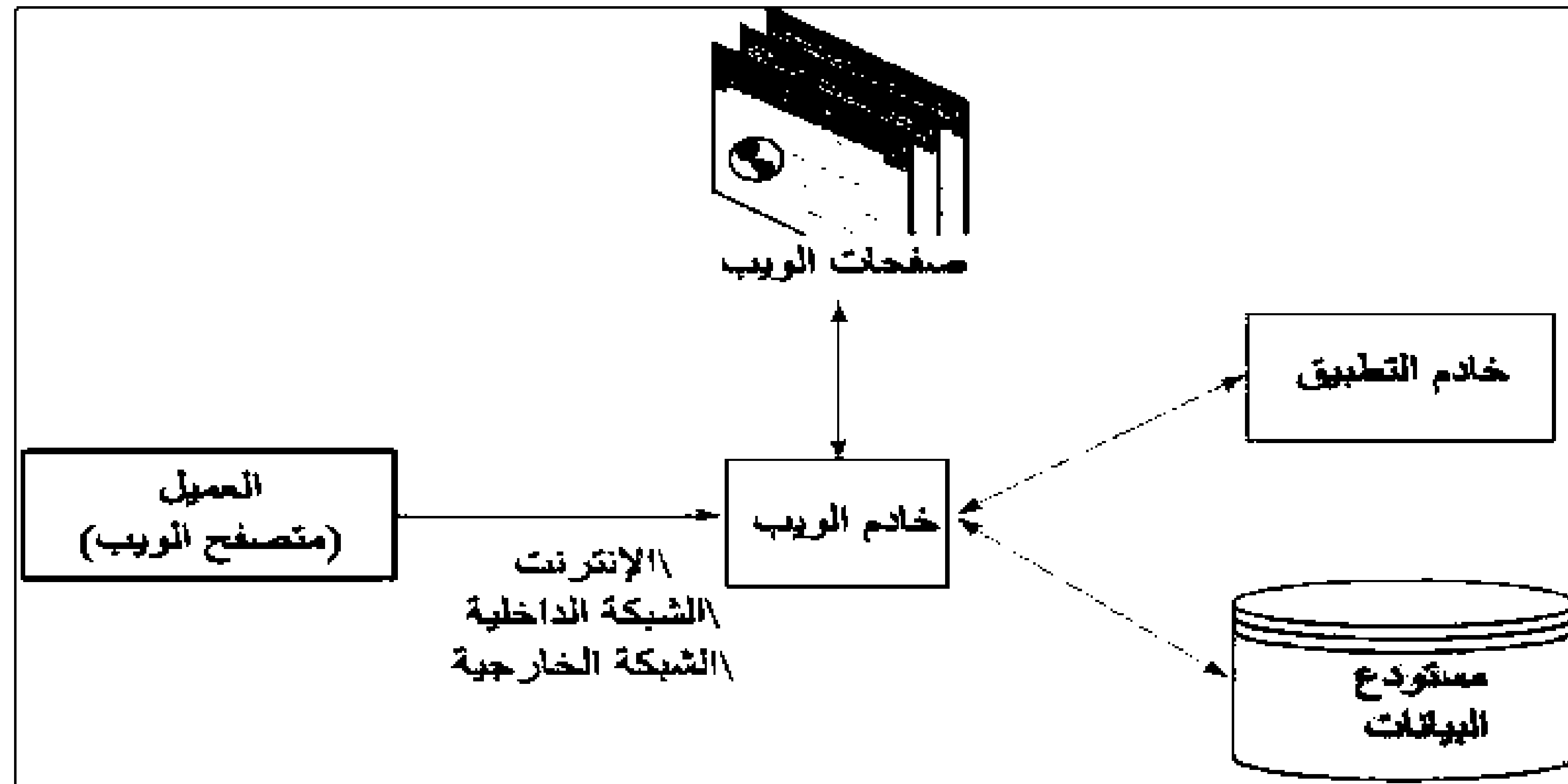
في بنية ثلاثية الطبقات؛ تحتوي الأنظمة التشغيلية على البيانات والبرمجيات؛ لحَصْر البيانات في طبقةٍ واحدةٍ (أيْ: الخادم)، ويكون مستودعُ البيانات في طبقة أخرى، في حين تحتوي الطبقة الثالثة على محرك DSS/BI/BA (أيْ: خادم التطبيق) والعميل (انظر: شكل ٣-٥). تتمُّ معالجة البيانات من المستودع مرتين، ويتمُّ إيداعها في قاعدة بيانات إضافية ومُتعدِّدة الأبعاد، تمَّ تنظيمها لسهولة عمل تحليل وعَرْض مُتعدِّد الأبعاد، أو نسخها في DMs. وتتمثل مَيْزة البنية الثلاثية في فصلها عن وظائف مستودع البيانات؛ مما يلغي قيود الموارد ويُسهِّل إنشاء DM.

وفي بُنيةٍ من مستويين، يعمل محرك DSS فعليًّا على نفس النظام الأساسي كمستودع للبيانات (انظر: شكل ٣-٦). ولذلك؛ فهي أكثر اقتصادًا من الهيكل الثلاثي. ويمكِن أن يكون للبنية ذات المستويين مشكلات في الأداء لمستودعات البيانات الكبيرة التي تعمل مع التطبيقات الكثيفة البيانات لدعم القرار.

يفترض جزءٌ كبيرٌ من الحكمة الشائعة اتباع نَهْجٍ استبدادي، معتبرًا أن أحد الحلول أفضل من الآخر، بصرف النظر عن ظروف المنظمة واحتياجاتها الفريدة. ولمزيدٍ من تعقيد هذه القرارات الهيكلية؛ يُركِّز العديد من الاستشاريين ومُطوِّري البرامج على جزءٍ واحد من البنية؛ مما يَحِدُّ من قدرتهم ودافعيتهم على مساعدة المنظمة من خلال الخيارات القائمة على احتياجاتها. ولكن يتمُّ التشكيك في هذه الجوانب وتحليلها. فعلى سبيل المثال: قدَّم Ball (2005) معايير اتخاذ القرار للمؤسسات التي تخطط لتنفيذ تطبيق ذكاء الأعمال وقد حدَّد بالفعل حاجته إلى DMs المتعددة الأبعاد؛ ولكنه يحتاج إلى المساعدة في تحديد البنية المتدرجة المناسبة. وتدور معاييره حول التنبؤ بالاحتياجات للمساحة وسرعة الوصول (للمزيد من التفاصيل انظر: Ball 2005).

يُعَدُّ كلٌّ من مستودعات البيانات والإنترنت من التقنيات الرئيسة التي تقدِّم حلولًا مهمةً لإدارة بيانات الشركة. وينتجُ عن دَمْج هاتين التقنيتين مستودعات البيانات على شبكة الإنترنت. ونعرض في الشكل (٣-٧) هيكل مستودعات البيانات على الويب. إذ تتكون البنية من ثلاث طبقات، وتضمُّ عميل الحاسب وخادم الويب وخادم التطبيقات. على جانب العميل؛ يحتاج المستخدِم إلى اتصال إنترنت ومتصفح ويب (يُفضَّل أن يكون Java) من خلال واجهة المستخدِم الرسومية المألوفة (GUI). ويُعَدُّ وسيط الاتصال بين العميل والخوادم هو الإنترنت/ الإنترانت/ الإكسترانت. بالنسبة للخادم؛ يتمُّ استخدامُ خادم ويب لإدارة تدفق المعلومات الداخلة والخارجة بين العميل والخادم. وتكون مدعومةً من قِبل كلٍّ من مستودع البيانات وخادم التطبيقات. كما يوفر مستودعات البيانات على شبكة الإنترنت العديد من المزايا الجذَّابة، بما في ذلك سهولة الوصول واستقلالية النظام وقلة التكلفة.

وتتشابه فنيات الويب الخاصة بمستودعات البيانات مع الفنيات الأخرى لمستودعات البيانات، والتي تتطلَّب اختيار تصميم لإسكان مستودع بيانات الويب مع خادم المعاملات أو كخادم منفصل/ خوادم منفصلة. وتُعَدُّ سرعة تنزيل الصفحات أحد الاعتبارات المُهِمّة في تصميم التطبيقات المستندة إلى الويب؛ ولذلك يجب تخطيط سعة الخادم بعناية.



شكل ٣-٧: بنية تخزين البيانات القائمة على الويب

يجبُ الوضع في الاعتبار العديدَ من المسائل عند تحديد أيِّ بنية سوف يتمُّ استخدامها. من بينها ما يلي:

- ما هو نظام إدارة قواعد البيانات (DBMS) الذي ينبغي استخدامه؟ يتمُّ إنشاء معظم مستودعات البيانات باستخدام RDBMS. وتُعَدُّ كلٌّ من Oracleمن شركة أوراكل (.oracle com) وSQL Server من شركة مايكروسوفت (microsoft.com/sql)، وDB2 من شركة IBM (http://www-01.ibm.com/software/data/db2)، هم الأكثر استخدامًا. يدعمُ كلٌّ من هذه المنتجات كلًا من العميل/ الخادم والهياكل المبنية على شبكة الإنترنت.
- هل ستتمُّ المعالجة على التوازي أم على التوالي؟ تُمكِّن المعالجة المتوازية العديد من وحدات المعالجة المركزية (CPU) من معالجة طلبات استعلام مستودع البيانات آنيًّا كما توفر القابلية للتوسُّع كلما لزم الأمر. يحتاج مُصمِّمو مستودع البيانات إلى تقرير ما إذا كان سيتمُّ تقسيم جداول قاعدة البيانات (أيْ: تقسيمها إلى جداول أصغر)؛ من أجل كفاءة الوصول، وتحديد المعايير التي ستُستَخدَم في المعالجة. وهذا هو الاعتبار المهمُّ الذي تقتضيه كميات كبيرة من البيانات الواردة في مستودع بيانات نموذجي. يمكن العثور على مَسْح حديث على مستودعات البيانات المتوازية والموزَّعة في Furtado (2009). لقد نجحت شركة تيراداتا (Teradata.com) في تبَنِّي هذا النهج بكفاءة، وكثيراً ما يتمُّ التوصية به.

- هل سيتمُّ استخدامُ أدوات ترحيل البيانات لتنزيل مستودع البيانات؟ إن نقل البيانات من نظامٍ قائم إلى مستودع للبيانات؛ هو مهمةٌ شاقةٌ ومُضنيةٌ. واعتمادًا على تنوُّع وموقع أصول البيانات؛ قد تكون الهجرة إجراءً بسيطًا نسبيًّا أو (على النقيض) مشروعًا يستغرق عدة أشهر. يجب استخدامُ نتائج تقييم شامل لأصول البيانات الموجودة لتحديد ما إذا كان سيتمُّ استخدامُ أدوات الترحيل أم لا، وإذا كانت الإجابة بنعم؛ فما هي القدرات المطلوبة في تلك الأدوات التجارية.
- ما الأدوات التي سيتمُّ استخدامها لدعم استرجاع البيانات وتحليلها؟ في كثيرٍ من الأحيان، يكون من الضروري استخدام أدوات مُتخصِّصة لتحديد مواقع البيانات الضرورية والوصول إليها وتحليلها واستخراجها وتحويلها وتنزيلها بشكلٍ دوري في مستودع البيانات؛ إذ يجب اتخاذ قرار بشأن اختيار أحد البدائل: ١- إما تطوير أدوات الترحيل داخليًّا، أو ٢- شراؤها من جهةٍ خارجية، أو ٣- استخدام الأدوات المتوفرة بنظام مستودع البيانات. وتستدعي عمليات الترحيل الفورية والمعقدة للغاية الاستعانة بأدوات ETL لجهات خارجية مُتخصِّصة.

بُنيات مستودعات البيانات البديلة:

على أعلى مستوى؛ يمكن تصنيفُ وجهات نظر تصميم هيكل مستودع البيانات في تصميم مستودع البيانات على مستوى المؤسسة (EDW)، وتصميم DM (Golfarelli & Rizzi, 2009). ونعرض في الشكل (٣-٨) a-e بعضَ البدائل لأنواع التصاميم الهيكلية الأساسية التي لا تكون (EDW) خالصةً ولا DM خالصة، ولكن تتنوع بين الهياكل الفنية التقليدية أو خارجها. وتشمل الهياكل الجديدة البارزة كلًّا من hub-and-spoke والهياكل الموحَّدة. تمَّ اقتراحُ الهياكل الخمسة المبيَّنة في الشكل (٣-٨) a-e بواسطة Ariyachandra وWatson (2005, 2006a,b). وفي دراسةٍ سابقة مستفيضة، حَدَّد كلٌّ من Sen وSinha (2005) خمس عشرة منهجية مختلفة لمستودعات البيانات. تُصنَّف مصادر هذه المنهجيات إلى ثلاث فئات عريضة: باعة التقنية الأساسية، وبائعي البنية التحتية، وشركات نمذجة المعلومات.

أ- **سجلات البيانات المستقلة:** يُمكِن القول: إنَّ هذا هو أبسط البدائل الهيكلية وأقلها تكلفةً. وقد تمَّ تطويرُ DMs للعمل مستقلةً بعضها عن بعض لتلبية احتياجات الوحدات التنظيمية المنفردة. وبسبب استقلاليتها، قد يكون لها تعريفات بيانات غير مُتسقة وأبعاد ومقاييس مختلفة؛ مما يؤدي إلى صعوبة تحليل البيانات عبر DMs (أيْ: إنه من الصعب، إن لم يكن من المستحيل، الحصول على «نسخة واحدة من الحقيقة»).

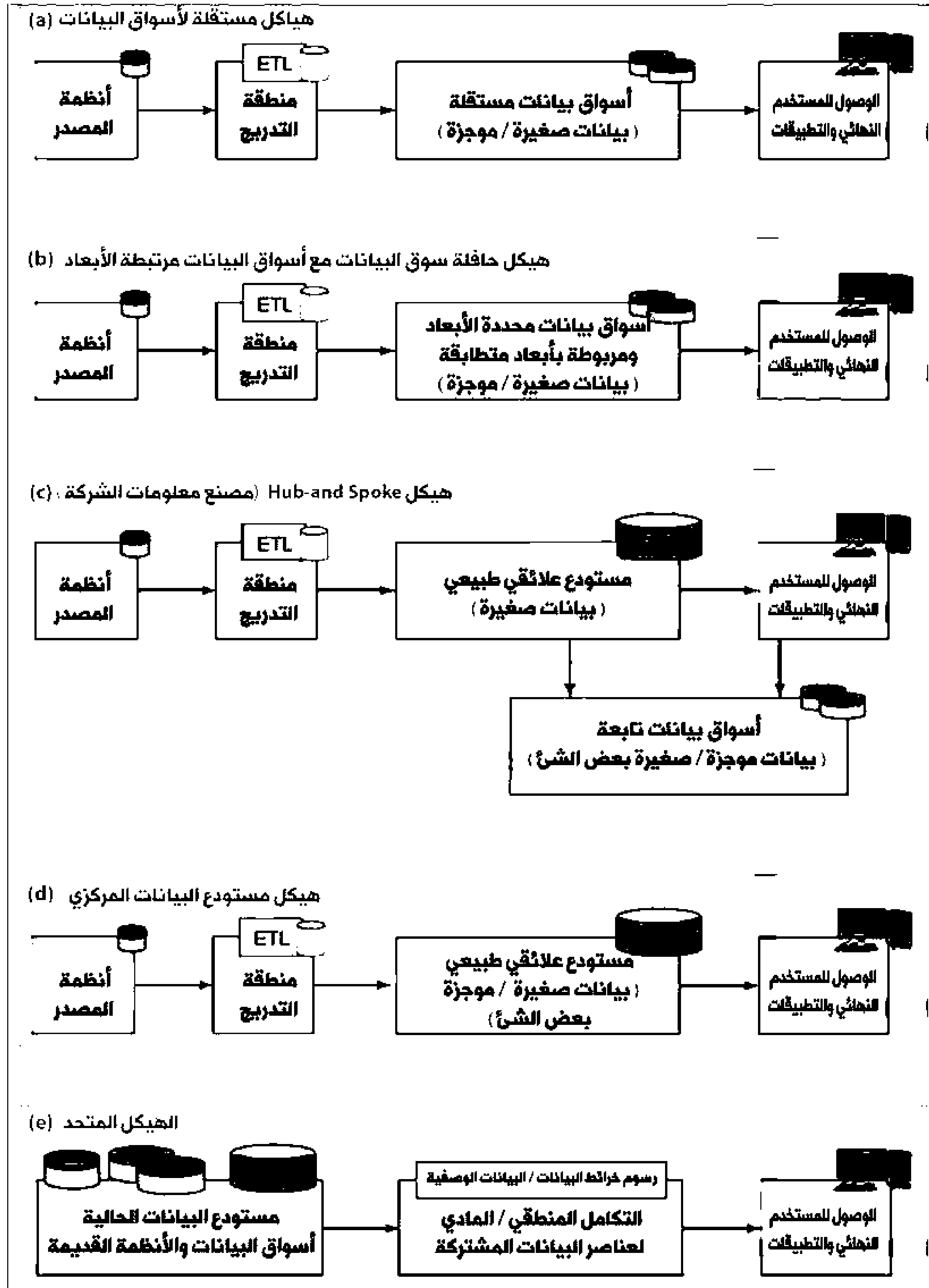
ب- **هيكل حافلة ســوق البيانات:** يُعَدُّ هذا الهيكل بديلًا قابلًا للتطبيق لـ DMs المســتقلة؛ إذ يتمُّ رَبْط الأسواق المنفردة بعضها مع بعض عن طريق نوعٍ من الوسائط. ونظرًا لأن البيانات تكون مرتبطةً بعضها ببعض؛ فهناك فرصةٌ أفضل للحفاظ على اتساق البيانات في جميع أنحاء المؤسسة (على الأقل عند مستوى البيانات الوصفية). وعلى الرغم من أن هذا الهيكل يسمح باستعلامات البيانات المعقدة عبر DMs؛ فإنَّ أداء هذه الأنواع من التحليل قد لا يكون بمستوى مُرضٍ.

ج- **هيــكل hub-and-spoke:** ولعلَّ هذا هو أشــهرُ هيكل لمســتودعات البيانات اليوم. وهنا ينصــبُّ الاهتمام على بناء بنية تحتية قابلة للتطوير وقابلــة للصيانة (غالبًا ما يتمُّ تطويرها بطريقة تكرارية، أيْ: موضوع تلو الآخر) والتي تتضمَّن مستودع بيانات مركزي والعديد من DMs (لكل وحدة تنظيمية). تســمح هذه البنية بســهولة التخصيص لواجهات المستخدِم والتقارير. وعلى الجانب السَّــلبي؛ تفتقر هذه البنية إلى الرؤية الشاملة للمؤسسة وقد تقود إلى تكرار البيانات، ووقت استجابة طويل للبيانات.

د- **مستودع البيانات المركزي:** يتشابه هيكلُ مستودع البيانات المركزي مع هيكل hub-and-spoke باســتثناء أنه لا تُوجَد DMs تابعة؛ ويُوجَد بدلًا منهــا EDW عملاقًا يلبِّي احتياجات جميع الوحــدات التنظيميــة. يوفِّر هذا النَّهج المركزي للمســتخدمين إمكانيــة الوصول إلى جميع البيانات الموجودة في مســتودع البيانات بدلًا من حصرهــا في DMs. وإضافةً إلى ذلك؛ فإنه يقلِّل من كمية البيانات التي يتعَيَّن على الفريق الفني نقلها أو تغييرها، وبالتالي تبسيط إدارة البيانات وقيادتها فنيًّا. وإذا تمَّ تصميمه وتنفيذه بشكلٍ صحيح؛ فإن هذا الهيكل يوفِّر رؤيةً شاملةً في الوقت المناسب عن المشروع لأيِّ شخص في أيِّ وقتٍ أينما كان داخل المنظمة.

هـ- **مســتودع البيانات الموحَّد:** يُعَدُّ النهجُ الموحَّد بمثابة امتيازٍ للقوى الطبيعية التي تفجِّر أفضل الخطــط لتطوير نظام مثالي. ويســتخدم جميع الوســائل الممكنة لدمج المــوارد التحليلية القادمة من مصادر مُتعدِّدة لمقابلة الاحتياجات المتغيرة أو ظروف العمل. وبصورة أساسية؛ ينطــوي النهج الموحَّد عــلى دَمْج أنظمة متباينة. وفي البنية الموحــدة؛ يتمُّ تَرْك هياكل دَعْم القــرار القائمة في مكانها، ويتمُّ الوصول إلى البيانات من تلك المصادر كلما لزم الأمر. كما يتمُّ دَعْــم النهج الموحَّد من قِبل مُورِّدي البرامج الوســيطة الذين يقترحون الاســتعلامات الموزَّعة والقدرات المشــتركة. تقدِّم هذه الأدوات القابلة للتجديد والقائمة على لغة الترميز (XML) للمســتخدمين رؤيةً شــاملةً لمصادر البيانات الموزَّعــة، بما في ذلك مســتودعات البيانات، و DMs والمواقع الإلكترونية والمستندات والأنظمة التشغيلية. عندما يختار المستخدِم كائنات الاســتعلام من هذا العرض، ويضغط زر الإرســال، تقوم الأداة بالاستعلام التلقائي عن المصادر الموزَّعة، وتضمُّ إليها النتائج، ومن ثمَّ تُقدِّمها للمســتخدِم. وبســبب مشكلات الأداء وجودة

البيانات؛ يتفق معظم الخبراء على أن الطرق المتبعة تعمل بشــكلٍ جيد لتكملة مستودعات البيانات وليس استبدالها (انظر: Eckerson, 2005).



شكل ٣-٨: هياكل بديلة لمستودع البيانات

وفي ٢٠٠٥ حــدَّد كلٌّ من Ariyachandra وWatson عشرة عوامل من المحتمل أن تؤثر على قرار اختيار الهيكل المستخدَم:

١- ترابط المعلومات بين الوحدات التنظيمية.

٢- احتياجات معلومات الإدارة العليا.

٣- الضرورة الملحَّة لمستودع البيانات.

٤- طبيعة مهامِّ المستخدِم النهائي.

٥- القيود على الموارد.

٦- عَرْض إستراتيجي لمخزن البيانات قبل التنفيذ.

٧- التوافق مع الأنظمة الحالية.

٨- القدرة الملموسة لفريق عمل تقنية المعلومات الداخلي.

٩- المشكلات الفنية.

١٠- العوامل الاجتماعية/ السياسية.

وتتشــابه هذه العوامل مع العديدِ من عوامل النجاح الموصوفة في الأدبيات الخاصة بمشــاريع نظم المعلومات ومشروعات DSS وذكاء الأعمال؛ إذ تُعَدُّ المسائل الفنية - بخلاف توفير التقنية التي تكون عمليًّا جاهزةً للاســتخدام - أمرًا مهمًّا؛ ولكنها غالبًا لا تكون مهمةً، مثل المشــكلات السلوكية، مثل: تلبية احتياجات المعلومات الخاصة بالإدارة العليا، ومشــاركة المســتخدِمين في عملية التطوير (عامل اجتماعي/ سياسي). ولكلِّ هيكل من هياكل مستودعات البيانات تطبيقات خاصة به متفاوتة الفاعلية، وبالتالي؛ توفر فوائد قصوى للمنظمة. ومع ذلك يبدو بشكلٍ عام أن هيكل DM هو الأقل فعاليةً في الاستخدام. وللمزيد من التفاصيل الإضافية انظر: Ariyachandra وWatson (2006a).

أيُّ البُنيات هي الأفضل؟

منذ أن أصبحت مســتودعات البيانات جزءًا مهمًّا من المؤسســات الحديثة، بات الســؤال عن بنية مســتودع البيانات الأفضل؛ هو موضوع المناقشــة الدائمــة. ويتزَعَّم هذه المناقشــات اثنان من زعماء مجال مســتودعات البيانات، هما: Bill Inmon وRalph Kimball. فدائمًا ما يدافع Inmon عن بنية hub-and-spoke (مثل: مصنع المعلومات المؤسسية)، في حين يُروِّج Kimball لحافلة هيكل DM ذات الأبعاد المطابقة. وهناك هياكلُ أخرى ممكنة؛ ولكن هذان الخياران هما نهجان مختلفان جوهريًّا، ولكلٍّ منهما مناصرون أقوياء. ولإلقاء الضوء على هذا الســؤال المثير للجدل، أجرى كلٌّ من Ariyachandra وWatson (2006a) دراســةً تجريبيةً. ولجمع البيانات قاما بإجراء اســتطلاع على الويب استهدفا فيه

الأفراد المشاركين في عمليات تنفيذ مستودع البيانات. تضمَّن الاستطلاعُ أسئلةً شخصيةً عن القائم بالإجابة عنه، وعن شركته، ومستودع بيانات الشركة، ومدى نجاح هيكل مستودع البيانات.

إجمالًا؛ قدَّم ٤٥٤ من المجيبين معلوماتٍ قابلةً للاستخدام. وتراوحت الشركاتُ التي تمَّ استطلاعها من صغيرة (ذات عائد أقل من ١٠ مليون دولار) إلى كبيرة (ذات عائدات تزيد على ١٠ مليار دولار). كانت معظم الشركات موجودةً في الولايات المتحدة بنسبة بلغت (٦٠٪) ومثلت مجموعةً متنوعةً من الصناعات، وشكَّلت صناعة الخدمات المالية معظم الاستجابات بنسبة بلغت (١٥٪). كانت البنية السائدة؛ هي بنية hub-and-spoke بنسبة (٣٩٪)، تليها بنية الحافلة (٢٦٪)، ثم البنية المركزية بنسبة (١٧٪)، ثم هيكل DMs المستقل بنسبة (١٢٪)، وجاء في المرتبة الأخيرة الهيكلُ المتحد بنسبة (٤٪). وكان النظام الأساسي الأكثر شيوعًا لاستضافة مستودعات البيانات هو Oracle بنسبة (٤١٪)، تليها مايكروسوفت (Microsoft) بنسبة (١٩٪)، ثم IBM بنسبة (١٨٪). وتراوح معدل (متوسط) الدخل الإجمالي من ٣٫٧ مليار دولار أمريكي لـ DMs المستقلة إلى ٦ مليار دولار للهيكل المتحد. وقد استخدموا أربعة مقاييس لتقييم نجاح الهياكل: (١) جودة المعلومات، (٢) جودة النظام، (٣) التأثيرات الفردية، و(٤) التأثيرات التنظيمية. واستخدمت الأسئلة مقياسًا من ٧ نقاط؛ إذ تشيرُ الدرجة الأعلى إلى بنية أكثر نجاحًا. يُوضِّح الجدول ٣-١ متوسط نتائج المقاييس لهياكل مستودعات البيانات.

	DMs المستقلة	هيكل الحافلة	هيكل hub-and-spoke	الهيكل المركزي (لا تعتمد على DMs)	الهيكل المتحد
جودة المعلومات	٤٫٤٢	٥٫٦١	٥٫٦٣	٥٫٣٢	٤٫٣٧
جودة النظام	٤٫٩٥	٥٫٠٦	٥٫٥٦	٥٫١٤	٤٫٩٦
التأثيرات الفردية	٥٫٠٨	٥٫٠٨	٥٫٢٦	٥٫٤٦	٥٫٥١
التأثيرات التنظيمية	٤٫٦٦	٥٫٤٣	٥٫٤٢	٥٫٠٣	٤٫٧٧

وكما تشيرُ نتائجُ الدراسة؛ سجَّلت DMs المستقلة أدنى الدرجات على جميع المقاييس. وتؤكِّد هذه النتيجة على الحِكمة التقليدية التي تقول بضعف الحلِّ الهيكلي المُقدَّم من DMs المستقلة. وجاء الهيكلُ الموحَّد في المركز قبل الأخير. وتمتلك الشركات في بعض الأحيان مِنصَّات دعم قرارات متباينة ناتجة عن عمليات الاندماج فيما بين الشركات، وقد تختار الطريقة المتحدة، على الأقل على المدى القصير؛ إذ تشيرُ النتائج إلى أنَّ الحلَّ المقدَّم من الهيكل المتحد ليس هو الأمثل على المدى الطويل. لكن الشيء المثير للاهتمام هو التشابه في المتوسطات بالنسبة لكلٍّ من الحافلة

وhub-and-spoke والهياكل المركزية. وتكون الاختلافات صغيرة بما فيه الكفاية؛ بحيث لا يمكن تقديمُ أيِّ مطالبات لتفوق هيكل معين على الهياكل الأخرى، على الأقل استنادًا إلى مقارنة بسيطة بين هذه المقاييس للنجاح. كما قاموا بجَمْع بيانات عن النطاق (والتي يتراوح من وحدة فرعية إلى نطاق الشركة) وحجم (أيْ: كمية البيانات المخزَّنة) المستودعات.

كما وجدوا أن بُنية hub-and-spoke تُستَخدَم عادةً مع تطبيقات أكثر على مستوى المؤسسة ومستودعات أكبر. وقاموا أيضًا بالتحقيق في التكلفة والوقت اللازمين لتنفيذ الهياكل المختلفة. بشكلٍ عام؛ كانت بنية hub-and-spoke هي الأكثر تكلفةً واستهلاكًا للوقت لتنفيذها.

أسئلة مراجعة على القسم ٣-٤:

١- ما هي أوجه التشابه والاختلاف الرئيسية بين هيكلٍ من طبقتين، وهيكلٍ ثلاثي الطبقات؟

٢- كيف أثر الويبُ على تصميم مستودع البيانات؟

٣- اذكر هياكل مستودعات البيانات البديلة التي تمَّت مناقشتها في هذا القسم.

٤- ما هي المسائل الواجب أخذها في الاعتبار عند تحديد البنية التي ستُستَخدَم في تطوير مستودع البيانات؟ اذكر أهمَّ ١٠ عوامل.

٥- أيٌّ من هياكل مستودعات البيانات هو الأفضل؟ ولماذا؟

٣-٥ تكامل البيانات وعمليات الاستخراج والتحويل والتنزيل (ETL):

إن الضغوط التنافسية العالمية، والطلب على عائد الاستثمار (ROI)، والإدارة والاستفسار عن المستثمرين واللوائح الحكومية؛ كلُّ هذه الأمور تجبر مديري الشركات على إعادة التفكير في كيفية دَمْج وإدارة أعمالهم، وعادةً ما يحتاج صانعُ القرار أن يصل إلى مصادر متعددة للبيانات، والتي يجبُ أن تكون متكاملةً. وقد كانت عمليةُ الوصول إلى مصادر البيانات عمليةً شاقةً وكبيرةً؛ وذلك قبل مستودعات البيانات، وDMs، وبرمجيات ذكاء الأعمال. وحتى مع أدوات إدارة البيانات الحديثة والتي تعتمدُ على شبكة الإنترنت؛ فإن التعرُّف على البيانات التي يمكن الوصول إليها وتقديمها إلى صانع القرار هي مهمةٌ غير ضرورية، غير أنها تتطلب متخصصين في قواعد البيانات ومع نموِّ حجم مستودعات البيانات تزداد مشكلات دمج البيانات هي الأخرى.

إنَّ تحليل الأعمال يحتاجُ إلى تطوُّر مستمر؛ إذ يُمكِن لعمليات الاندماج والاستحواذ والمتطلبات التنظيمية وإدخال قنوات جديدة أن تؤدي إلى تغييرات في متطلبات ذكاء الأعمال. وإضافةً إلى البيانات التاريخية والمنظفة والموحَّدة والمتحصَّل عليها في الوقت المناسب؛ فإنَّ مُستخدِمي الأعمال

يحرصون على الوصول إلى البيانات غير المهيكلة في الوقت المناسب بدون تأخير. ويجبُ أن يتكامل كلُّ شيءٍ مـع محتويات مسـتودع البيانات الموجود. علاوةً على ذلك؛ أصبـح الوصولُ عبر أجهزة المسـاعد الرقمي الشخصي (PDA) ومن خلال التعرُّف على الكلام والتوليف أكثر شيوعًا؛ مما يزيد من تعقيد قضايا التكامل (Edwards, 2003). وتتضَمَّن العديدُ من مشـاريع التكامل أنظمةً على مستوى المؤسسة. وقد قدَّم Orovic في ٢٠٠٣م قائمةً مرجعيةً بما يصلح وما لا يصلح عند محاولة القيـام بمشروع من هذا القبيل. ومن الصعـب دَمْجُ البيانات التي تأتي من قواعد بيانات مختلفة وغيرها من المصادر المتباينة؛ خاصةً عندما لا يتمُّ ذلك بشـكلٍ صحيح؛ إذ إنه من الممكن أن يؤدي إلى كارثة في أنظمة واسعة النطاق مثل ERP، CRM، ومشاريع سلاسل التوريد (Nash, 2002).

تكامل البيانات:

يشتملُ تكاملُ البيانات على ثلاث عمليات رئيسية؛ هذه العمليات عندما يتمُّ تنفيذها بشكلٍ صحيح؛ فإنها تسـمح بالوصول إلى البيانات وإتاحتها لمجموعة من ETL، وأدوات التحليل، وبيئة مسـتودعات البيانات، وهذه العمليات هي: الوصـول إلى البيانات: وتعني القدرة على الوصول إلى البيانات واستخراجها من أيِّ مصدر بيانات، اتحاد البيانات؛ أيْ: تضمين آراء الأعمال في مخزن البيانـات المتعددة، تغيـير الالتقاط؛ والذي يتم بناءً على تحديد التغيـيرات التي تمَّ إدخالها على مصادر بيانات المؤسسـة. انظر الحالة العملية ٣-٢ للاطلاع على مثالٍ لكيفية الاسـتفادة من BP Lubricant في تنفيذ مسـتودع بيانات يدمج البيانات من عِدة مصادر. فقد قام بعضُ المورِّدين، مثل SAS Institute, Inc بتطوير أدوات تكامل البيانات بشـكلٍ قوي. ويتضمَّن مشروع تكامل البيانات SAS للمؤسسـة أدواتِ تكامل بيانات العملاء التي تعمل على تحسين جودة البيانات في عملية التكامل. كما يساعد Oracle Business Intelligence Suite أيضًا على تكامل البيانات.

ويتمثل الغرض الرئيسي من مسـتودع البيانات في دَمْـج البيانات من أنظمة مُتعدِّدة؛ بحيث تتيح تقنيات التكامل المختلفة دمج البيانات والبيانات الوصفية:

- تكامل تطبيقات المؤسسات (EAI).
- العمارة الموجهة نحو الخدمة (SOA).
- تكامل معلومات المؤسسة (EII).
- الاستخراج والتحويل والتحميل (ETL).

يوفِّر تكاملُ تطبيقات المؤسسات (EAI) وسيلة لدفع البيانات من أنظمة المصدر إلى مستودع البيانـات. وهـو ينطوي على دَمْج وظائـف التطبيقات ويركِّز على مشـاركة الوظائف (بدلًا من

البيانات) عبر الأنظمة، وبالتالي تمكين المرونة وإعادة الاستخدام. وعلى نحو تقليدي؛ فقد ركَّزت حلول تكامل تطبيقات المؤسسات (EAI) على تمكين إعادة استخدام التطبيق على مستوى واجهة برمجة التطبيقات. وفي الآونة الأخيرة؛ يتمُّ إنجاز تكامل تطبيقات المؤسسات (EAI) باستخدام خدمات (SOA) (والتي هي مجموعة من العمليات أو المهام التجارية) التي تم تعريفها وتوثيقها بشكلٍ جيد. ويُعَدُّ استخدام خدمات شبكة الإنترنت طريقةً مُتخصِّصةً لتنفيذ (SOA). كما يُمكِن استخدامُ (EAI) لتسهيل الحصول على البيانات مباشرةً في مستودع بيانات شِبه فوري أو لتوصيل القرارات إلى أنظمة OLTP، وهناك العديدُ من الطرق والأدوات المختلفة لتنفيذ (EAI).

ويُعَدُّ تكامل معلومات المؤسسات (EII) مجالًا جيدًا باعتباره أداةً متطورةً تقومُ بدَمْج البيانات الواردة من مجموعة متنوعة من المصادر في الوقت المُحدَّد، مثل: قواعد البيانات المترابطة، وخدمات شبكة الإنترنت، وقواعد البيانات المتعددة الأبعاد؛ بالإضافة إلى كونه آليةً لسَحْب البيانات من أنظمة المصدر لتلبية طلب الحصول على المعلومات. وتستخدم أدوات (EII) بيانات التعريف المعرَّفة مسبقًا لتجميع وجهات النظر التي تجعل البيانات المدمجة (المتكاملة) تبدو مرتبطةً بالمستخدمين النهائيين، وقد يكون XML هو الجانب الأكثر أهميةً في EII؛ وذلك لأن XML يسمح بوَضْع علامات على البيانات إما في وقت الإنشاء أو في وقت لاحق، وهذه العلامات يمكن توسيعها وتعديلها بما يسمح باستيعاب أي مساحة من المعرفة (انظر: Kay, 2005).

ولطالما كان تكامل البيانات المادية؛ هو الآلية الرئيسة لصُنْع نظرة متكاملة مع مستودعات البيانات وDMs. ومع ظهور أدوات تكامل معلومات المؤسسة (EII) (انظر: Kay, 2005)؛ فإن أنماط تكامل البيانات الافتراضية الجديدة أصبحت مُجدِيةً وقابلةً للتطبيق، وقد ناقش كلٌّ من Manglik و(2005) Mehra فوائد وقيود أنماط تكامل البيانات الجديدة التي يمكن أن توسِّع المنهجيات المادية التقليدية لتقديم رؤية شاملة للمشروع.

حالة عملية ٣-٢

بريتش بتروليوم BP لمواد التشحيم تحقق نجاح BigS

قامت شركة بريتش بتروليوم لزيوت التشحيم بتأسيس برنامج BigS لمتابعة نشاط الدَّمج الأخير لتوفير معلومات إدارة متسقة وشفافة على مستوى العالم؛ إضافةً إلى تقديم ذكاء الأعمال، وBigS في الوقت المناسب والذي يوفر عرضًا تفصيليًّا ومتكاملًا للأداء عبر وظائف، مثل: التمويل، والتسويق والمبيعات، والإمدادات وخدمات التخطيط والتنفيذ.

وتُعَدُّ BP واحدةً من أكبر مجموعات النفط والبتروكيماويات في العالم؛ إذ تُعَدُّ جزءًا من مجموعة BP plc؛ فشركة BP لمواد التشــحيم هي شركةٌ رائدةٌ في ســوق السيارات العالمية، وتشــتهر الشركة بعلامة زيوت كاسترول (Castrol) التجارية، كما تعمل في أكثر من مائــة دولة ولديها عشرة آلاف موظف. ومن الناحية الإســتراتيجية؛ فإنَّ شركة BP لمواد التشــحيم تركِّز على تحســين خدمة عملائها، وزيادة فاعليتها في أسواق السيارات. وفي أعقاب نشاط الاندماج مؤخرًا تمرُّ الشركة بعملية تحوُّل لتصبح أكثر فاعليةً ومرونةً، واغتناماً لفرص النمو السريع.

التحَدِّي:

عقب نشــاط الاندماج الأخير أرادت شركة BP لمواد التشــحيم تحســينَ الاتســاق والشــفافية، وســهولة الحصول على معلومات الإدارة وذكاء الأعمال. وللقيام بذلك كان من الضروري دَمْج البيانات المُحتفَظ بها في أنظمة مصدر معدل التوزيع دون التأخير في إدخال نظام مُوحَّد لتخطيط موارد المؤسسات.

الحل:

نفَّــذَت شركة BP لمواد التشــحيم برنامــج «ذكاء الأعمال التجاريــة والمعايير العالمية» (BigS)، وهي مبادرةٌ إستراتيجيةٌ لإدارة وذكاء الأعمال، وفي لُبِّ برنامج BigS يُوجَد ما يُسمَّى Kalido وهو حل EDW التكيُّفي لإعداد وتنفيذ وتشغيل وإدارة مستودعات البيانات.

وقد أسهم حلُّ EDW المتكامل من Kalido في دَعْم تكامل البيانات المعقدة للبرنامج التجريبــي ومتطلبات تقديم التقارير المتنوعة، وللتكيُّــف مع متطلبات إعداد التقارير المتطوِّرة للبرنامج؛ فقد تمكَّن البرنامج أيضًا من تعديل بنية المعلومات الأساسية بسهولة عاليــة مع الحِفاظ على جميع المعلومات؛ إذ يقوم النظام بدَمْج وتخزين المعلومات من أنظمة مصادر متعددة لتوفير طرق عرض مُوحَّدة من أجل:

- تســويق العائــدات والهوامش الخاصة بالعميل بالنســبة لقطاعات الســوق مع التعمق في تفاصيل مستوى الصوت.

- زيادة التقارير الخاصة بفواتير المبيعات والتكاليف الجمركية التفصيلية والرواتب الفعلية.

- تقديم بيانــات الأرباح والخســائر والميزانية العمومية وبيانــات التدفق النقدي الموحَّــدة عــلى الصعيد العالمي، مــع التدقيق في مراجعة حســابات العمــلاء والديون وخدمات التخطيط والتنفيذ؛ بحيث يتمُّ معالجة الحركة وعَرْضها من خلال نظام موحَّد عبر مِنصات ERP متعددة في نظام تخطيط موارد المؤسسات.

الفوائد:

توفِّر BigS المعلومات اللازمة لمســاعدة الأعمال التجارية على تحديد عددٍ وافر من الفرص التجارية؛ من خلال تحســين رؤية البيانات المتسقة في الوقت المناسب؛ من أجل تحقيق أقصى قدرٍ من فرص العمل لزيادة الهوامش أو إدارة التكاليف المرتبطة بها.

وتتضمَّن الاستجابات النموذجية لفوائد البيانات المتسقة الناتجة عن BigS ما يلي:

- تحسين التناسق والشفافية في بيانات الأعمال.
- إعداد تقارير أسهل وأسرع وأكثر مرونة.
- التوافق بين المعايير العالمية والمحلية على حَدٍّ سواء.
- دورة تنفيذ سريعة وفعَّالة؛ من حيث التكلفة بالإضافة إلى مرونتها.
- أقل قدرٍ من الاضطراب للعمليات التجارية القائمة والأعمال اليومية.
- تحديد مشكلات جودة البيانات والتشجيع على حَلِّها.
- تحسين القدرة على الاستجابة بذكاء للفرص التجارية الجديدة.

أسئلة للمناقشة:

١- ما هو BigS؟

٢- ما هي التحَدِّيات، وما هو الحل المقترح والنتائج التي تمَّ الحصول عليها مع BigS؟

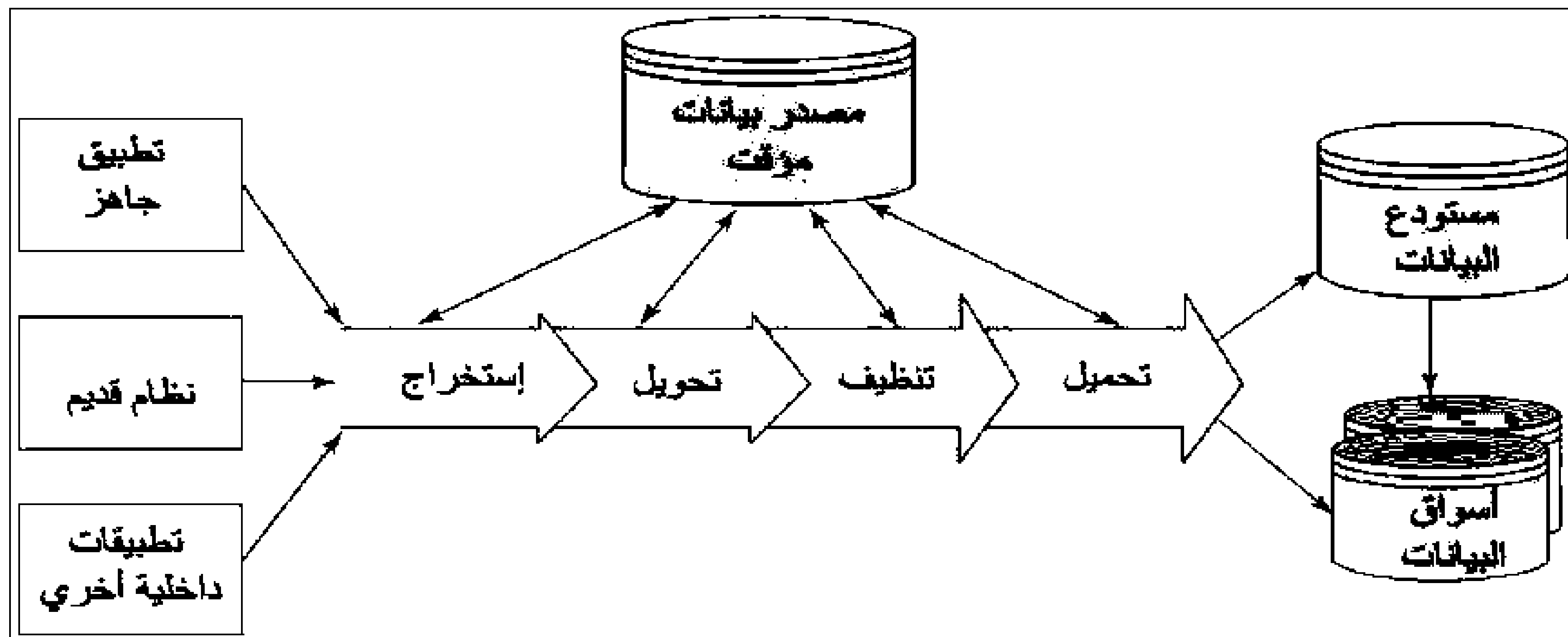
Sources: Kalido. BP Lubricants. http://kalido.com/download/BP-Lubricants.pdf (accessed July 2016); BP Lubricants, www.bp.com/en/global/corporate/about-bp/bp-at-a-glance.html (accessed July 2016).

الاستخراج، والتحويل، والتحميل:

يتضمَّن الجانبُ التقني من عملية مســتودعات البيانات استخراج وتحويل وتحميل (ETL). وتقوم تقنية (ETL) - والتي كانت موجودةً منذ فترة قصيرة - بدور رئيس في العملية واســتخدام مستودعات البيانات. وتُعَدُّ عملية استخراج وتحويل وتنزيل البيانات (ETL) عنصراً أساسياً في أيِّ مشروع بيانات مركزية. وغالبًا ما يواجه مديرو تقنية المعلومات تحَدِّيات؛ لأن عملية اســتخراج وتحويل وتنزيل البيانات عادةً ما تستهلك ٧٠٪ من الوقت في مشروع بيانات مركزية.

تتكوَّن عملية (ETL) من: الاســتخراج (ويعني: قراءة البيانات من قاعدة بيانات واحدة أو أكثر)، والتحويل (أيْ: تحويل البيانات المُســتَخرَجة من شــكلها السابق إلى الشــكل الذي تحتاجُ إليه؛ بحيث يمكِن وضعه في مستودع بيانات أو ببساطة قاعدة بيانات أخرى)؛ إضافةً إلى التنزيل (أيْ: وَضْع البيانات في مســتودع البيانــات). ويحدث التحويلُ باســتخدام القواعد أو جداول البحــث أو عن طريق دَمْج البيانــات ببيانات أخــرى. ويتمُّ دَمْج وظائف قواعد البيانات الثلاثة في أداةٍ واحدة لســحب البيانات من واحدة أو أكثر من قواعد البيانات، ووَضْعها في قاعدة بيانات أخرى مُوحَّدة، أو مستودع بيانات.

وتقــوم أدوات (ETL) أيضًــا بنقل البيانــات بين المصادر والأهداف وتوثيــق كيفية تغيُّر عناصر البيانات (كالبيانات الوصفية) في أثناء انتقالها بين المصدر والهدف؛ إضافةً إلى تبادل البيانات الوصفية مــع التطبيقات الأخرى حســب الحاجة، وإدارة جميع عمليات التشــغيل والعمليات الأخرى (مثل: الجدولة، وإدارة الأخطاء، وســجلات التدقيق، والإحصائيات). وتُعَــدُّ عملية (ETL) عمليةً في غاية الأهميــة لتكامل البيانات، وكذلك لمســتودعات البيانــات. والغرض من هــذه العملية، هو تنزيل المســتودع ببيانات مُدمَجة ومنتقاة، ومن الممكن أن تأتي البيانات المُستَخدَمة في عمليات(ETL) من أيِّ مصــدرٍ كـ: تطبيق حاســب مركزي، أو تطبيــق (ERP)، أو أداة CRM، أو ملف ثابت، أو جدول بيانات Excel، أو حتى قائمة انتظار رسائل. وفي الشكل ٣-٩ نحَدِّد الخطوط العريضة لعملية (ELT).



شكل ٣-٩: عمليات ETL

وتنطــوي عملية ترحيل البيانات إلى مســتودع البيانات على التنقيــب في البيانات من جميع المصــادر ذات الصِّلــة. وقد تتكــوَّن مصادر البيانات من ملفات مُســتخلَصَة مــن قواعد بيانات OLTP، وجداول البيانات، وقواعد البيانات الشخصية (على سبيل المثال: Microsoft Access)، أو الملفات الخارجية. وعادةً ما تتمُّ كتابة كافة ملفات الإدخال في مجموعةٍ من الجداول المؤقتة،

والتـــي يتمُّ تصميمها من أجل تســـهيل عملية التنزيل. ويحتوي مســـتودع البيانات على العديد من قواعد العمل التي تحَدِّد أشـــياء، مثل: كيفية اســـتخدام البيانات، وقواعد التلخيص، وتوحيد السِّمات المشفرة، وقواعد الحساب. ويجب تصحيح البيانات المتعَلِّقة بملفات المصدر قبل تنزيلها في مســـتودع البيانات. ومن فوائد مســـتودع البيانات المصمَّم تصميمًا جيدًا هو أن هذه القواعد يمكن تخزينها في مســـتودع البيانات الوصفية، وتطبيقها على مســـتودع البيانات بشـــكلٍ مركزي. وهـــذا يختلف عـــن طريقة OLTP، التي تحتوي عـــادةً على قواعد البيانـــات والأعمال التجارية المنتشرة في جميع أنحاء النظام. ويمكن إجراء عملية تنزيل البيانات في مستودع البيانات إمَّا من خلال أدوات تحويل البيانات التي توفر واجهة مُســـتخدِم رســـومية للمساعدة في تطوير وصيانة قواعد الأعمال التجارية، أو من خلال طرق أكثر تقليدية، مثل تطوير البرامج أو الأدوات المساعدة لتنزيـــل مســـتودع البيانات؛ وذلك باســـتخدام لغات البرمجة، مثل: لغـــات PL/SQL أو C++ أو Java أو دوت نت Framework. وهذا القرار ليس ســـهلًا بالنســـبة للمنظمات؛ إذ تؤثر العديدُ من المشـــكلات على قرار المؤسسة بشراء أدوات تحويل البيانات، أو إنشاء عملية التحويل نفسها، وتتمثل هذه المشكلات في:

- أدوات تحويل البيانات الغالية الثمن.
- قد يكون لأدوات تحويل البيانات منحنى تعلُّم طويل.
- من الصعب قياس أداء منظمة تقنية المعلومات؛ حتى تتعلَّم المنظمة استخدام أدوات تحويل البيانات.

وعلى المدى الطويل؛ ينبغي تبسيط نهج صيانة أداة تحويل مستودع البيانات للمنظمة. ومن الممكن أن تكون أدوات التحويل أيضًا فعَّالة في الكشف والتنظيف (بمعنى إزالة أية حالات شاذة في البيانات). وتعتمد OLAP وأدوات التنقيب في البيانات على مدى جودة تحويل البيانات.

وكمثالٍ على (ETL) الفعَّالة؛ فإن شركة Motorola, Inc تستخدم (ETL) لتغذية مستودعات البيانـــات الخاصة بهـــا؛ إذ تقوم شركـــة Motorola بتجميع المعلومات من ثلاثـــين نظامًا مختلفًا للمشـــتريات، ومن ثم تقوم بإرســـالها إلى مســـتودع بيانات SCM العالمـــي لتحليل إجمالي إنفاق الشركة. (انظر: Songini, 2004).

وقـــد صنَّـــف Solomon (2005) تقنيات ETL إلى أربـــع فئات، وهي: متطـــورة، وممكنة، وبســـيطة، وبدائية. ومن المســـلَّم به عمومًا أن أدوات الفئة المتطورة ســـتؤدي إلى توثيق عملية ETL بشكلٍ أفضل وأكثر دقة مع تطوُّر مشروع مستودع البيانات.

وعلى الرغم من أنه من الممكن للمبرمجين تطوير برنامج من أجل ETL؛ فإنه من الأسهل استخدام أداة ETL القائمة (الموجودة بالفعل). ونعرض فيما يلي لبعض المعايير المهمة في اختيار أداة ETL (انظر: Brown, 2004):

- القدرة على القراءة والكتابة إلى عددٍ غير محدودٍ من أساليب بناء مصادر البيانات.
- التقاط تلقائي للبيانات الوصفية وتسليمها.
- تاريخ المطابقة للمعايير المفتوحة.
- واجهة سهلة الاستخدام للمطوِّر والمستخدِم الوظيفي.

قد يكون أداء ETL الشامل علامةً على سوء إدارة البيانات، ونقص أساسي في إستراتيجية إدارة بيانات متماسكة. وقد أشار Karacsony (2006) إلى وجود علاقة مباشرة بين حجم البيانات المتكررة وعدد عمليات ETL. فعندما يتم إدارة البيانات بشكلٍ صحيح كأصل لمؤسسة؛ فإن جهود ETL تنخفض بشكلٍ كبير، كما يتمُّ التخلُّص من البيانات المتكررة بالكامل. مما يؤدي إلى توفير كبير في الصيانة وزيادة الكفاءة في التطوير الجديد مع تحسين جودة البيانات. إنَّ سوء تصميم عمليات ETL يجعل تكلفة صيانتها وتغييرها وتحديثها باهظة. وبناءً على ذلك؛ فإنه من الأهمية بمكان اتخاذ الخيارات المناسبة من حيث التقنية والأدوات لاستخدامها في تطوير وصيانة عملية ETL.

ويتوفر عددٌ من أدوات ETL المجمَّعة؛ إذ يقوم مُورِّدو قواعد البيانات حاليًّا بتوفير قدرات ETL التي تعمل على تحسين أدوات ETL المستقلة والتنافس معها. وتقرُّ SAS بأهمية جودة البيانات وتقدِّم أول حلٍّ متكاملٍ تمامًا في هذه الصناعة يدمج ETL، وبيانات نوعية لتحويل البيانات إلى قيمة الأصول الإستراتيجية، ومن بين مُزوِّدي برامج ETL الآخرين شركات مايكروسوفت (Microsoft)، وOracle، وIBM، وInformatica، وEmbarcadero، وTibco. وللحصول على معلومات إضافية حول ETL انظر: Golfarelli و Rizzi (2009)، وKaracsony (2006)، وSongini (2004).

أسئلة مراجعة القسم ٣-٥:

١- صِف تكامل البيانات.

٢- صِف الخطوات الثلاث لعملية ETL.

٣- لماذا تُعَدُّ عملية ETL مهمةً جدًا لجهود مستودعات البيانات؟

٣-٦ تطويرُ مستودع البيانات:

يُعَدُّ مشروع مســتودعات البيانات مهمةً كبيرةً لأيِّ مؤسســة، وهو أكــثرُ تعقيدًا من مشروع اختيار وتنفيذ حاســب مركزي بســيط؛ نظرًا لأنه يضمُّ العديدَ من الأقسام والعديدَ من واجهات الإدخال والإخراج بالإضافة إلى تأثيره عليها، كما أنه من الممكن أن يكون جزءًا من إستراتيجية عمل CRM. ويوفِّر مســتودعُ البيانات العديدَ من المزايا، والتي يمكن تصنيفها إلى مزايا مباشرة ومزايا غير مباشرة. وتتضمَّن المزايا المباشرة ما يلي:

- يمكِن للمستخدِمين النهائيين إجراءُ تحليل مستفيض بطرقٍ متعددة.
- يُمكِن توحيدُ وجهات النظر فيما يتعلَّق ببيانات الشركة (أيْ: إصدار صيغة واحدة للحقيقة).
- يُمكِن الحصولُ على معلومات أفضل وفي الوقت المناســب؛ إذ يســمحُ مستودع البيانات بمعالجة المعلومات؛ ليتمَّ إعفاؤها من أنظمة التشغيل المكلِّفة إلى خوادم منخفضة التكلفة. ولذلك؛ فإنه من الممكن معالجة الكثير من طلبات الحصول على معلومات للمستخدِم الأخير بسرعة أكبر.
- يُمكِن أن يؤدي إلى مستوى أداء أفضل؛ إذ يُحرِّر مستودع البيانات عملية الإنتاج؛ نظرًا لأن بعض متطلبات تقارير نظام التشغيل يتمُّ نقلها إلى DSS.
- تبسيط الوصول إلى البيانات.

وتنشــأ المزايا غير المباشرة نتيجةً لاستخدام المستخدِمين النهائيين للمزايا المباشرة. وبشكلٍ عام؛ فإنَّ هذه المزايا تعَزِّز المعرفة بالأعمال التجارية، وتقدِّم فائدةً تنافســية؛ إضافةً إلى تحســين خدمة العملاء وإرضائهم، كما أنها تُسهِّل اتخاذ القرار، وتساعد في إصلاح العمليات التجارية. وبالتالي؛ فإنَّ مستودعات البيانات تُسهِم بشكلٍ قوي لإضفاء مَيْزة تنافسية للمؤسسات (Parzinger & Frolick, 2001)، وللاطلاع على مناقشة مُفصَّلة حول كيفية حصول المؤسسات على مستويات استثنائية من الأربــاح، انظر: Watson وGoodhue و(2002) Wixom. ونظرًا للفوائد المُحتمَلة التي يمكن أن يوفرها مســتودع البيانات والاســتثمارات الكبيرة في الوقت والمال التي يتطلبها هذا المشروع؛ فإنه من الأهمية بمكان أن تقوم المؤسسة بتنظيم مشروع مستودع بياناتها لتحقيق أقصى قدرٍ من فرص النجاح. وإضافةً إلى ذلك يجبُ على المؤسســة بطبيعة الحال أن تأخذ في الاعتبار التكاليف. وقد وَصَف Kelly (2001) منهج عائد الاســتثمار (ROI) والذي يأخــذ في الاعتبار مزايا فئات عمليات الحِفــظ (أيْ: الأموال التي يتمُّ توفيرها من خلال تحســين وظائف دَعْم القرار التقليدية)، والجمع (أيْ: الأموال التي يتمُّ توفيرها بســبب جمع المعلومات ونشرها آليًّا)، والمســتخدمين (أيْ: الأموال المحفوظة أو المُكتسَــبة من القرارات التي يتمُّ اتخاذها باســتخدام مســتودع البيانات). وتشمل

التكاليف كلّ ما يتعلق بالأجهزة، والبرمجيات، والنطاق التردُّدي للشبكة، والتنمية الداخلية؛ إضافةً إلى الدَّعم الداخلي، والتدريب، والاستشــارات الخارجية. ويتمُّ حساب القيمة الحالية الصافية على مدى العمر المتوقع لمســتودع البيانات. ونظرًا لأن المزايا تُقسَّــم تقريبًا إلى: ٢٠٪ لعمليات الحِفظ، و٣٠٪ للجمع، و٥٠٪ للمســتخدمين؛ فقد أشــار Kelly إلى أنه يجبُ إشراك المستخدمين في عملية التطوير، والذي يُعَدُّ عامل نجاح حاسم للأنظمة التي تُعنَى بالتغيير في المنظمة.

وتُقدِّم الحالة العملية ٣-٣ مثالًا لمخزن البيانات الذي تمَّ تطويره، وتنفيذه على أساس تنافس شديد لشركة مبيعات تجزئة هولندية، وقد أدَّى الجَمْع بين قدرات مســتودعات تيراداتا التخزينية والقدرات التحليلية مع وجود بنية تحتية للمؤسسة باستخدام حلول SAP إلى تحقيق نجاح هائل في السوق.

حالة عملية ٣-٣

استخدام تحليلات تيراداتا لحلول SAP تُسرِّع عملية تسليم البيانات الكبيرة

إنَّ الشركة التي تمَّ تصنيفها ضِمن هذه الدراســة الخاصة باقتناء المنتج، هي عبارة عــن شركةٍ مملوكة ملكية خاصة لتاجر تجزئــة هولندي، وتتكون من ٢٨٠٠ متجر، هذا التاجر لديه ١٥ علامة تجارية مختلفة عبر العديد من الدول الأوروبية. تشمل منتجات عديــدة من ألعاب الأطفــال إلى أدوات الطهي. وكلُّ علامة تجارية لهــا البنية التحتية الخاصة بها. وتتم إدارة كلِّ كيان تجاري بشــكلٍ مستقل عن باقي الـ ١٥ شركة الأخرى؛ وذلــك في طريقة تطوير عملياته، والحِفاظ على أنظمة تراثه، واتخاذ القرارات التجارية الخاصة بالتمويل وتقنية المعلومات، وسلسلة التوريد والعمليات العامة.

الخلفية:

ومــن أجل تلبية احتياجات بيئة تنافســية دائمــة التطوُّر؛ فإنَّ ذلك يتطلَّــب رؤيةً عالميةً للأعــمال؛ وهو الأمر الذي يمثل تحديًّا لمتاجر التجزئة الكبيرة لهذا التاجر الهولندي مع قيام ١٥ علامة تجارية مســتقلة بإدارتها. وللحصول على رؤية أفضل وزيادة كفاءة الأعمال وانخفاض التكاليف؛ قرَّر بائع التجزئة وَضْع إستراتيجية مشتركة لإدارة البيانات في نظامٍ مركزي باستخدام قســم تقنية معلومات واحد. وتعني مركزية البيانات أنَّ جميع العلامات التجارية سوف تتمُّ إدارتها في مســتودع بيانات واحد، وتنفيذها حســب العلامة التجارية، مــع مراعاة العمليات التجارية الفردية واحتياجات كلِّ كيان مستقل. ويتمثل أكبرُ تحدٍّ يواجه متاجر التجزئة الكبيرة في ضرورة دَمْــج العديدِ من الأنظمة، بما في ذلك أنظمة (SAP ERP) الخاصة بالـ (١٥) علامة تجارية، وأنظمة إدارة المستودعات، وأنظمة نقطة البيع (POS)، والبيانات الرئيسة للمواد.

ومــع التركيز على الحِفاظ على مرونة الحركة التجارية للمبيعات وتحليل الهامش؛ فقد كان هــدف تاجــر التجزئة هو توفير الوصول إلى مســتوى معاملات البيانات الناشــئة ممَّا يقرب من ٥٠ جدول SAP داخل كلِّ نظام ERP. وقد زاد الأمرُ تعقيدًا خاصةً فيما يتعلق بالانتقال إلى نَهْج مركزي؛ بســبب الفروق الدقيقــة عبر نظام تخطيط موارد الخمس عشرة علامة تجارية. وقد احتاج العملُ على إدخال هذه البيانات في المســتودع المركزي إلى ٤٠٠ يــوم من الجهد لكلِّ نظام. وبالتالي؛ فقد احتاجوا إلى طريقةٍ لتبرير نفقاتهم، وتطوير قيمة مقترحة مستمرة لنهج مستودع البيانات الخاصة بهم، وتطوير طريقةٍ لتسهيل هذه العملية.

قصة اكتساب المنتج:

ومــن الناحية التجارية؛ فإن التركيز ينصبُّ على إنشــاء مِنصَّــة تحليلية مركزية مع إمكانية الوصول إلى رؤية شــاملة لبيانات المعاملات. ونظرًا للطبيعة الموسمية للتجزئة؛ فــإن القدرة على الاســتفادة مــن بيانات عدة ســنوات أمرٌ مهمٌّ للمســاعدة في تحديد الاتجاهات الموســمية، وإنشــاء التوقعات، وتطوير الأسعار والعروض الترويجية. ويكون الهدف من كلِّ ذلك هو تحسين الرؤية وتوفير حرية التحليلات عبر سلسلة التوريد والمواد والمبيعات والتســويق لمساعدة هذه المؤسسة لتصبح أكثر كفاءةً في طريقة عملها. وبناءً على ذلك؛ فقد قــام بائع التجزئة باختيار قاعدة بيانات تيراداتا؛ نظرًا لأنه يمكنه التعامل مــع تحليلات المعاملات بالإضافة إلى توفــير قدرات تحليلات متقدِّمة. وكان هدفهم هو دَعْم التحليلات التشــغيلية والمرونة عن طريق تنزيل البيانات دون تطوير نظام DMS أو نماذج منطقية أخرى قبل طرح المستخدمين أسئلة الأعمال. وقد مكَّنَهم هذا الأسلوب من حِفظ البيانات مركزيًّا داخل قاعدة بيانات تيراداتا مع توفير المرونة المســتقبلية التي تتعلق بالوصول إلى البيانات والتقارير والتحليلات لجميع العلامات التجارية.

وفي سبيل التقليل من تعقيدات SAP ERP؛ قضَت الشركة الأشهر الستة الأولى تعمل بكامل جهدها؛ من أجل تكامل SAP للعملاء المحليين. وبعد ســتة أشــهر من العرض أدركــوا المخاطر، وأوقفوا المشروع للتحقيق فيما إذا كانــت هناك طرق أفضل للتعامل مــع هذه المشــكلة. ولأول مرة تقوم الشركة باســتخدام نظــام SAP SI والذي احتاج تقريبًــا لـ ٤٠٠ يوم حتى يتمَّ فقط تنزيــلُ البيانات من أول نظام لـ SAP ERP. وهذا الأمــرُ لن يوفِّــرَ قيمةً بالسرعة الكافية؛ لذلك فقد قام تاجــر التجزئة الهولندي بتصعيد المشكلة بتصعيد المشكلة ومحاولة استكشاف حل جديد يعمل بطريقة أوتوماتيكية في حــلِّ تيراداتا جديد يعمل بطريقة أوتوماتيكية في عمليــات الحصول على البيانات عند

اســتخدام SAP ERP. وقد وقع الاختيار على تحليلات تيراداتا لحلول SAP؛ نظرًا لأنه تــمَّ تصميمه خصيصًا لمعالجة التحَدِّيــات المرتبطة بإحضار البيانات من SAP ERP إلى قاعــدة بيانات تيراداتا. كما يوفر الحــلُّ نهجًا آليًّا (تلقائيًّا) لدَمْج بيانات SAP ERP في مســتودع البيانات كما مكَّنَهم من تنزيل البيانات المطلوبــة للعلامة التجارية الأولى في غضون ٥ أيام فقط بدلًا من ٤٠٠ يوم المقدَّرة باستخدام النظام السابق. وقد قضى بائع التجزئة ٤٥ يومًا إضافيًّا لإضافة ٢٥ جدولًا مخصصًا (Z) وإعداد البيانات للاستهلاك. مما أدى إلى تسريع دَمْج بيانات SAP بنسبة ٨٠٠٪، وبالتالي توفير ٣٥٠ يومًا من العمل.

التحَدِّيات:

يــؤدي الجَمْع بين مشروع توحيــد تخطيط موارد المؤسســات بالكامل عبر العديد من الأنظمة القديمة إلى إنشــاء مشروع به العديدُ من التعقيدات. وعلى الرغم من أن تحليلات تيراداتا لحلول SAP قد أتاح التشــغيل التلقائي لجزء إدارة البيانات المرتبطة بـ SAP مــن المشروع؛ فإنَّ تاجر التجزئة لا يزال يواجه تحديات فنية؛ بســبب حقيقة أن مبادرة مســتودعات البيانات الخاصة به قد تمَّ دمجُها مع مشروع تكامل أوسع. وكان النهج الذي اتبعوه هو توحيد الأدوات ووَضْع إطار عمل مع أول علامتين تجاريتين يمكِن تطبيقهما على الطرح التدريجي لبقية المنظمة.

في البداية كانوا بحاجةٍ إلى التوحيد القياسي على أداة ETL وتطوير منهجية جديدة وطريقة الاستفادة من ETL. وقد استخدموا أداة ETL كأداة تحويل (Extract ELT) للتنزيــل والحِفاظ على تكامــل بيانات المعاملات المعقدة. وقــد انتهي هذا التاجرُ إلى اختيــار ®Informatica كمعيــار ETL وبيئــة ETL الخاصة به باســتخدام أداة ELT كمُحرِّك للبيانات وجدولة المهمة فقط.

ثانيــاً، بالإضافة إلى تخزيــن بيانات المعاملات الدقيقة؛ فقد اســتطاع تاجر التجزئة الاســتفادة من مِنصَّة تيراداتــا لتنفيذ جميع تحويلات الأعمال في قاعــدة البيانات عند نقل البيانات إلى بيئة إعداد التقارير. وقد سمح لهم هذا النهج بالاحتفاظ بنسخة من المعاملات الدقيقة، والاســتفادة من عمليــات التكامل خارج المربع المتوفرة في تحليلات تيراداتا لحلول SAP؛ بغرض إضافة سياق إلى بيانات SAP، وتسخير قوة قواعد البيانات لتطبيق التحويلات والتحليلات الأخرى.

ثالثاً، كان أمرُ الحصول على بيانات عالية الجودة أمرًا ضروريًّا بالنسبة لهم. كما أنهم يريدون التأكُّد من أن البيانات يُمكِن الوصول إليها وإدارتها على نحو مُتســق. وتوضِّح

الأرقام المادية أهمية إدارة البيانات بالنسبة إلى هذا التاجر. وقد تمَّ تنظيمُ أرقام المواد بشــكلٍ مختلفٍ عبر أنظمــة متعددة، وكان من الممكن توفيقهــا خلال عملية التنزيل/ النمــوذج. وقد مكَّنهم هذا الهيــكل الجديد من التغلُّب على التحَــدِّي الذي واجههم بسهولة من خلال إنشاء طرق عَرْض المواد الفريدة في مستودع البيانات لتنسيق الأرقام المادية لإعداد التقارير.

وأخــيرًا؛ فقد كان الأمرُ يتطلَّــبُ طريقةً ذكيةً لتقديم البيانــات التحليلية والوصول للتقارير المتخصصة، والتي يمكنها أيضًا أن تلبِّي المتطلبات المتنوعة للعلامة التجارية.

ومن خلال الاستفادة من شراكات تيراداتا مع موفِّري الحلول مثل MicroStrategy®؛ فقد تمكَّن هذا التاجرُ من الوصول إلى البيانات الدقيقة المخزَّنة في مســتودع البيانات في أثناء استخدام أدوات ذكاء الأعمال لتطبيق الخوارزميات المناسبة، والاستفادة من المرونة المُصمَّمة في حَلِّ مستودع البيانات.

وفي البدايــة؛ كان تطويــرُ مســتودع البيانات باعتبــاره محوراً مركزيًّــا للوصول إلى البيانات؛ أمرًا صعبًا بســبب الحاجة إلى تطوير إطــار عمل جديد ومنحنى التعلُّم العام بســبب التغيير في أسلوب تصميم مستودع البيانات. ولحُسْن الحَظِّ؛ فإنه بمجرد تطوير هذا الإطار، كان التكاملُ باســتخدام برنامج تحليلات تيراداتا لحلول SAP بسيطًا وقابلًا للتكــرار. ووفقاً لرؤيــة مخطط المشروع لدى متاجر التجزئــة الأوروبية: «يُعَدُّ برنامج تحليــلات تيراداتا لحلول SAP حلاً متكاملاً وسريعــاً ومرناً، كما أنه يحقق تنميةً أسرع، ويقلل نســبة المخاطرة، بالإضافة إلى كونه نموذجاً دلاليًّا متكاملاً، ويوفر الوصول المباشر إلى البيانات التفصيلية».

الدروس المستفادة:

وبوجــهٍ عام؛ فإن هدف تاجر التجزئة هو توفير إســتراتيجية تنفيــذ قابلة للتكرار؛ مــن خلال العلامات التجارية الخاصة بالشركة؛ وذلك من أجل تمكين أفضل القرارات في العمل؛ إضافةً إلى تحســين كفاءة الأعمال، وخفض تكاليف التشــغيل من خلال مركزية تقنية المعلومات. وعلى الرغم من أنهم ما زالوا في المراحل الأولى من المشروع؛ فإنهم قد تعلَّموا من تنفيذ عملية دَمْج علامتهم التجارية الأولى في مستودع بيانات تيراداتا. ونظرًا لاســتخدام صاحب الشركة لبرنامج تحليلات تيراداتا لحلول SAP؛ فقد تمكَّنوا من تسريع وقت تقييم أنشــطة التكامل وتبســيطها. إضافةً إلى تمكُّنهم مــن تطوير بعض الفوائد التالية لتطبيقها على دَمْج علاماتهم التجارية اللاحقة والمشاريع المماثلة.

وتتمثل هذه الفوائد فيما يلي:

- أَخْــذ الوقت الــكافي في العناية الواجبة والتعرُّف على التقنيــات/ الحلول المتوفرة لدعم عمليات التنفيذ. وفي هذه الحالة تمكن صاحب الشركة من الاستفادة من برنامج تحليلات تيراداتا لحلول SAP؛ مما أدى إلى تقليل وقت التقييم وتمكينه من التركيز على التحليلات بدلًا من التكامل.

- تطوير إطار عملٍ لتمكين العمليات المتكرِّرة التي يُمكِن أن تعالج تعقيدات الكمية الهائلة من البيانات والاحتياجات المخصَّصة للأعمال.

- الحِفاظ قدر الإمكان على بساطة تصميم النظام؛ لضمان اعتماد التقنية والنشاط التجاري.

- التأكُّد من توافق القرارات الفنية مع الرؤية الشاملة لتمكين سرعة الأعمال التجارية.

- تطويــر نهجٍ قياسيٍّ لإدارة البيانات؛ لضمان ســلامة البيانــات التي تمتدُّ إلى ما بعد عملية التنفيذ؛ بحيث يفهم مســتخدمو الأعمال والمســتخدمون الفنيون كيفية تطبيق البيانات للتقارير والتحليلات.

- تحديد مهلة للتأخير؛ لضمان وَضْع حلول -لكلٍّ من مستودعات البيانات والتكامل- تدعم الاحتياجات. مما يعني ضمان أنَّ حلَّ Teradata SAP يدعم أيضًا احتياجاتهم التشغيلية.

وتُطبَّــق هــذه الدروس المســتفادة للتنفيذ على نطاقٍ أوســع، واســتخدام برنامج تحليلات تيراداتا لحلــول SAP. وقد التزم صاحبُ الشركة بمركزية بنيته التحتية وإدارة علاماته التجارية بشــكلٍ أكثر فعالية. مما أدَّى إلى الاســتفادة من هذه الطريقة لجعل العمليــة تتمُّ بطريقة أوتوماتيكية مع تقليل الوقت اللازم لإتمامها؛ بســبب القدرة على إرساء حلٍّ مُستَهدَف لربط حلول تخطيط موارد المؤسسات لتحليلاتهم.

أسئلة للمناقشة:

١- ما هي التحَدِّيات التي واجهتها متاجرُ التجزئة الهولندية الكبيرة؟

٢- ما هو الحلُّ المقترح متعدد الأطراف؟ وما هي التحَدِّيات التي واجهت التنفيذ؟

٣- ما هي الدروس المستفادة؟

Source: Teradata case study. (2015). 800 percent: Use of Teradata® Analytics for SAP® Solutions accelerates Big Data delivery. assets.teradata.com/resourceCenter/downloads/CaseStudies/EB8559_TAS_Case_Study.pdf?processed=1 (accessed July

2016); Enterprise Management, Teradata-SAP Solution to Big Data analytics. www.enterprisemanagement.com/research/asset.php/3047800-/Percent:-Use-of-Teradata-Analytics-for-SAP-Solutions-Accelerates- Big-Data-Delivery (accessed July 2016).

ومــن الواضــح أن تحديدَ هدف العمــل، واجتماع المســتخدِمين النهائيين لــلإدارة على دَعْم المــشروع، بالإضافة إلى تحديد الأُطُر الزمنيــة والميزانيات المعقولــة، وإدارة التوقعات، كلها أمورٌ بالغة الأهمية لنجاح مشروع مســتودعات البيانات؛ إذ إنَّ إســتراتيجية مستودعات البيانات هي مخططٌ للتطبيق الناجح في إدخال مســتودع البيانات. ويجب أن تقوم هذه الإستراتيجية بوَصْف الأهداف التي تسعى الشركة إلى تحقيقها، ولماذا تريدُ تحقيق هذه الأهداف، وماذا ستفعل عندما تحقــق هذه الأهداف. كما ينبغي أن تضع في اعتبارها رؤية المنظمة وهيكلها وثقافتها. ولمعرفة الخطوات التي يُمكِن أن تســاعد في وَضْع إســتراتيجية مرنة وفعَّالة يمكنك الاطلاع على (Matney, 2003). عند وَضْع خطة ودعم مستودع البيانات؛ فإن المؤسسة تحتاج إلى بحث مُورِّدي مستودع البيانات، (راجع الجدول ٣-٢ للحصول على قائمة عينات من المورِّدين، ويمكنك أيضًا مراجعة معهد مســتودعات البيانات [twdi.org] وبناة المعلومات [informationbuilders.com]). كما يُقدِّم العديد من المورِّدين برامج تجريبية لمستودعات البيانات الخاصة بهم ومنتجات ذكاء الأعمال.

جدول ٣-٢: قائمة عيِّنات لمورِّدي مستودعات البيانات

بائع	عروض المنتجات
كائنات الأعمال (businessobjects.com)	مجموعة شاملة من برمجيات ذكاء الأعمال وبرامج تصوير البيانات (المملوكة حاليًّا لـSAP)
شركاء الحاسب (cai.com)	مجموعة شــاملة مــن أدوات ومنتجات مســتودع البيانات (DW)
مرآة البيانات (datamirror.com)	قسم، وإدارة، وثمرة أداء DW
مجموعة مزايا البيانات (dataadvantagegroup.com)	برنامج البيانات الوصفية
Dell (dell.com)	خوادم DW
تقنيات Embarcadero (embarcadero.com)	قسم، وإدارة، وثمرة أداء DW
Greenplum (greenplum.com)	مزود حلول البيانات، وموفر حلول البيانات (مملوك الآن من قِبل EMC)
Harte-Hanks (harte-hanks.com)	ثمرة وخدمات إدارة علاقات العملاء (CRM)

بائع	عروض المنتجات
HP (hp.com)	خوادم DW
Hummingbird Ltd. (hummingbird.com)	محركات DM ومستودعات الاستكشاف
حلول Hyperion (hyperion.com)	مجموعة شـاملة من أدوات، ومنتجات، وتطبيقات DW
IBM InfoSphere (www-01.ibm.com/software/data/infosphere)	تكامل البيانات، وDM، وإدارة البيانات الرئيسـية، ومنتجات البيانات الكبيرة
Informatica (informatica.com)	قسم، وإدارة، وثمرة أداء DW
مايكروسوفت (microsoft.com)	أدوات ومنتجات DW
Netezza	برنامج وموفر الأجهزة DW (أجهزة DW) (مملوكة الآن لشركة IMB)
Oracle (وتشمل: PeopleSoft وoracle.com :Siebel)	أدوات ومنتجات وتطبيقات DW وERP وCRM
معهد SAS (sas.com)	أدوات ومنتجات وتطبيقات DW
Siemens (siemens.com)	خوادم DW
Sybase (sybase.com)	مجموعة شاملة من أدوات وتطبيقات DW
تيراداتا (Teradata.com)	أدوات DW، وأجهزة DW، واستشارات وتطبيقات DW

مناهج تطوير مستودع البيانات:

تحتاج العديدُ من المؤسسات إلى إنشاء مستودعات البيانات المُستَخدَمة لدعم القرار، وهناك نهجان يستخدمان في هذا الغرض، وهما:

- **النهج الأول:** وهو نَهْج ذكاء الأعمال Inmon ll والذي يُطلَق عليه «أب (Father) مستودعات البيانات»؛ إذ يؤيد Inmon نهج التطوير من أعلى إلى أسـفل والذي يتكيَّف مع أدوات قاعدة البيانات التقليدية الارتباطية؛ وذلك لتلبية احتياجات تطوير مسـتودع البيانات على مسـتوى المؤسسة، ويُعرَف هذا النهج أيضًا باسم نهج EDW.
- **النهــج الثاني:** هو نَهْج Ralph Kimball والذي اقترح نهجًا تصاعديًّا من القاعدة إلى القمة؛ إذ يستخدم نمذجة الأبعاد، ويُعرف هذا النهج أيضًا باسم نهج DM.

إنَّ معرفة أوجه التشـابه والاختلاف بين هذين النهجين يسـاعدنا على فَهْم مفاهيم مستودع البيانات الأساسـية (وللمزيـد انظر: Breslin, 2004). وفيما يلي جـدول ٣-٣ والذي يقارن بين هذين النهجين، ويقدِّم وصفًا تفصيليًّا لهما:

جدول ٣-٣: التباين بين نهجي التطوير DM وEDW

الجهود	نهج DM	نهج EDW
الإطار	المجال موضوع واحد	المجالات عدة مواضيع
وقت التطوير	شهور	سنة
تكلفة التطوير	١٠,٠٠٠ دولار إلى ١٠٠,٠٠٠ دولار	١,٠٠٠,٠٠٠,٠٠٠ +
صعوبة التطوير	منخفضة إلى متوسطة	عالية
المتطلبات المسبقة لتبادل البيانات	عام (داخل منطقة العمل)	عام (عبر المؤسسة)
المصادر	فقط بعض الأنظمة التشغيلية والخارجية	العديد من الأنظمة التشغيلية والخارجية
الأحجام	ميجابايتس إلى عدة جيجابايتس	جيجابايتس إلى بيتابايتس
أفق زمني	بالقرب من البيانات الحالية والتاريخية	البيانات التاريخية
تحويلات البيانات	منخفض إلى متوسط	متوسط
تحديث التردد	كل ساعة، يوميًا، أسبوعيًا	أسبوعيًا، شهريًا
التقنية		
المعدات	محطات العمل والخوادم والإدارات	خوادم المؤسسة وأجهزة الحاسب المركزية
نظام التشغيل	ويندوز ولينكس	يونيكس، Z/OS, OS/390
قواعد البيانات	مجموعة العمل أو خوادم قاعدة البيانات القياسية	خوادم قواعد بيانات المؤسسة
الاستعمال		
عدد المستخدمين المتزامنين	10s	من 100sإلى 1000s

الجهود	نهج DM	نهج EDW
أنواع المستخدمين	محللي منطقة الأعمال والمديرين	محللي الشركات وكبار المديرين التنفيذين
إلقاء الضوء على الأعمال	تحسين الأنشطة داخل مجال الأعمال	التحسين الانسيابي واتخاذ القرارات

المصادر: مقتبس من Van den Hoven, J. (2003). بيانات المخططات: خطة كبيرة، وبناء صغير. كتيب في إدارة IS، الطبعة الثامنة، بوكا راتون، فلوريدا: CRC Press، Ariyachandra, T., & Watson, H.2.

- **نموذج INMON** (طريقة EDW): ويركِّز نهجُ Inmon على التطوير من القمة إلى القاعدة، باستخدام منهجيات وأدوات تطوير قاعدة بيانات ثابتة، مثل: الرسوم البيانية للعلاقات بين الكيانات (ERD)، وتعديل نهج التطوير الحلزوني. ولا يمنع نهج EDW إنشاء DMs. ويُعَدُّ DMW نظاماً مثاليًّا في هذا النهج؛ لأنه يوفر وجهة نظر متسقة وشاملة للمؤسسة، وقد قدم Murtaza (1998) إطار عمل لتطوير EDW.
- **نموذج KIMBALL:** إنَّ نموذج إستراتيجية DM لـ Kimball هو «خطةٌ كبيرةٌ، وبناءٌ صغير». و DM هو مستودع البيانات الموجَّهة بالموضوع أو الموجَّهة بحسب القسم. فهو نسخة مُصغَّرة من مستودع البيانات الذي يركِّز على طلبات قسم معين، مثل: التسويق أو المبيعات. ويطبِّق هذا النموذج نمذجة بيانات الأبعاد، والتي تبدأ بالجداول. وقد دعا Kimball إلى منهجية تطوير تستلزم اتباع نهج تصاعديٍّ (من القاعدة إلى القمة)، وهو ما يعني في حالة مخازن البيانات بناء DM واحد في كلِّ مرة.

ما هو النموذج الأفضل؟

لا تُوجَد إستراتيجية واحدة تناسب الجميع لمستودعات البيانات. فمن الممكن أن تتطوَّر إستراتيجية تخزين بيانات المؤسسة من DM بسيط إلى مستودع بيانات مُعقَّد استجابةً لمطالب المستخدِمين، ومتطلبات أعمال المؤسسة، ونضج المؤسسة في إدارة موارد البيانات الخاصة بها. وبالنسبة للعديدِ من الشركات؛ فإن نموذج DM في كثيرٍ من الأحيان هو خطوة أولى مريحة لاكتساب الخبرة في بناء وإدارة مستودع بيانات في أثناء عرض مُستخدِمي الأعمال بالإضافة إلى فوائد الوصول بطريقة أفضل إلى بياناتهم، وتشير علامة DM عادةً إلى قيمة الأعمال الخاصة بمستودعات البيانات. وفي النهاية؛ تُعَدُّ هندسة EDW والتي تدمج DMs القديمة ومستودعات البيانات هي الحل المثالي (انظر: الحالة العملية ٣-٤). ومع ذلك؛ فإنَّ تطوير نماذج DMs الفردية يمكنها أن توفر فوائد عديدة في طريقها نحو تطوير EDW؛ خاصةً إذا كانت المؤسسة غير قادرةٍ

أو ليس لديها رغبة في الاســتثمار في مشروع على نطاق واســع. كما يمكن أن تظهر دلائل جدوى ونجاح نموذج DMs في تقديم المزايا. وقد يؤدي ذلك إلى اســتثمار في EDW. ويُلخِّص الجدول ٣-٤ أهمَّ الاختلافات المميزة بين النموذجين.

جدول ٣-٤: الفروق الأساسية بين نهجي Inmon وKimball

الصفة المميِّزة	Inmon	Kimball
المنهجية وأسلوب البناء		
النهج العام	من أعلى إلى أسفل (تنازلي)	من أسفل إلى أعلى (تصاعدي)
هيكل البناء	مستودع بيانات «يغذي» قواعد البيانات الإدارية على مستوى المؤسسة (صغير)	نموذج DMs عملية تجارية واحدة، ويتحقق الاتساق المؤسسي؛ من خلال ناقل البيانات ومطابقة الأبعاد
تعقد الأسلوب	معقدة جدًا	بسيطة إلى حَدٍّ كبير
مقارنة مع منهجيات التنمية القائمة	مشتقة من المنهجية الحلزونية	عملية من أربع خطوات؛ خروجًا عن أساليب RDBMS
مناقشة التصميم المادي	دقيق إلى حَدٍّ ما	بسيط نسبيًا
نمذجة البيانات		
اتجاه البيانات	موضوع، أو بيانات مستندة	عملية موجهة
الأدوات	تقليدية (رسوم بيانية لعلاقة الكيان [ERD]، ورسوم بيانية لتدفق البيانات [BDF])	نمذجة الأبعاد، خروجًا عن النمذجة العلائقية
إمكانية وصول المستخدم النهائي	منخفض	متوسط
الفلسفة		
الجمهور الرئيسي	مُتخصِّصو تقنية المعلومات	المستخدمون النهائيون
الموقع في المؤسسة	جزءٌ لا يتجزأ من قاعدة معلومات الشركات	محوِّل وخادم البيانات التشغيلية

الصفة المميَّزة	Inmon	Kimball
الموضوع	تقديـم حَـلٍّ تقنيٍّ سـليم يعتمـدُ على أساليب وتقنيات قاعدة البيانات المجرَّبة	تقديم حلٍّ يُسهِّل على المستخدمين النهائيين الاسـتعلام عـن البيانـات بشـكلٍ مبـاشر والحصول عليها في أوقات استجابة معقولة

Sources: Adapted from Breslin, M. (2004, Winter). Data warehousing battle of the giants: Comparing the basics of Kimball and Inmon models. Business Intelligence Journal, 9(1), 6-20; Ariyachandra, T., & Watson, H. (2006b). Which data warehouse architecture is most successful? Business Intelligence Journal, 11(1).

اعتبارات إضافية لتطوير مستودع البيانات:

تريد بعضُ المنظمات الاسـتعانة بمصادر خارجية بالكامل في جهود مستودعات البيانات؛ إذ إنهم لا يريـدون اقتناء البرمجيات والأجهزة، كما أنهم لا يريدون التعامل مع إدارة نظم المعلومات الخاصة بهم، وليس لديهم سوى بديلٍ واحد وهو استخدام مستودعات البيانات المستضافة. وفي هذه الحالة؛ فـإن الحل المثالي هـو قيام شركة أخرى لديها الكثـيرُ من الخبرة والمهارة، بتطوير مسـتودع البيانات والحِفاظ عليه. غير أنَّ هناك مخاوف تتعلق بالأمان والخصوصية مع استخدام هذا النهج (وللحصول على المزيد من التفاصيل راجع الرؤية الفنية ٣-١).

تمثيل البيانات في مستودع البيانات:

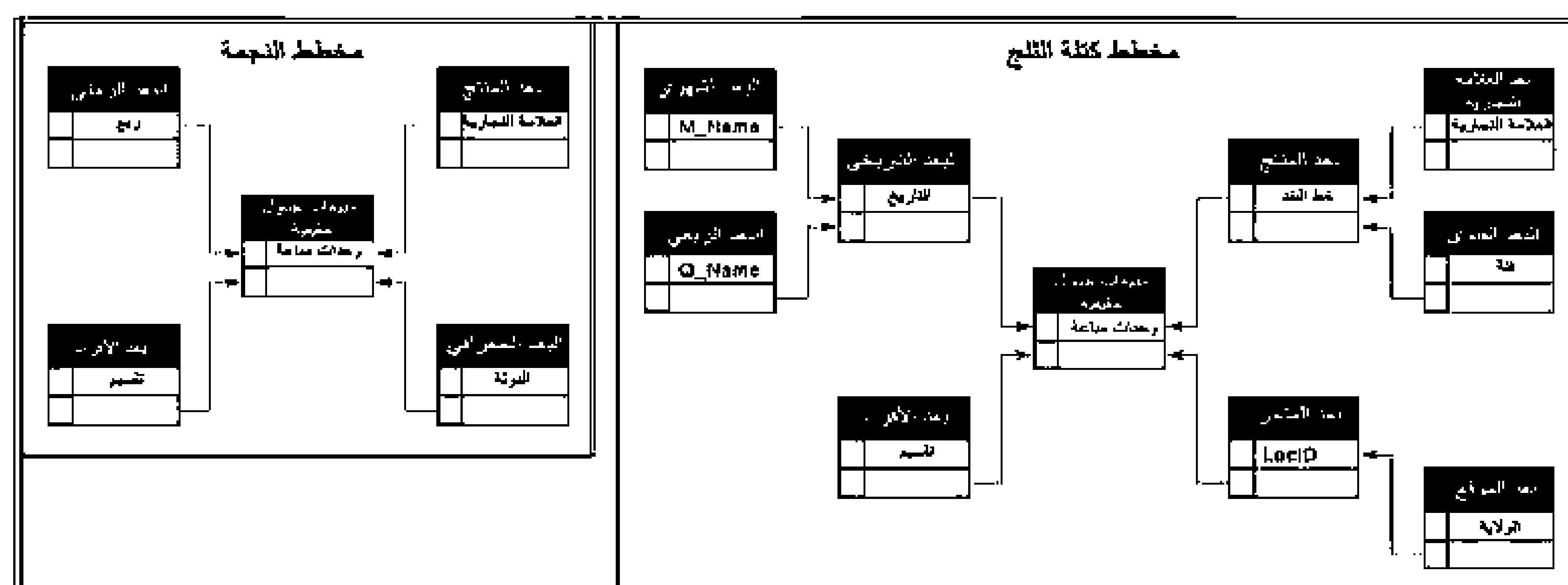
يُظهِر هيكلُ مسـتودع البيانات النموذجي في الشـكل ٣-٤ إمكانيةَ إجراء العديد من أشكال بنية مسـتودع البيانات (انظر: الشكل ٣-٨). وبغضِّ النظر عن البنية الهندسية؛ فقد كان تصميمُ تمثيل البيانات في مسـتودع البيانات يعتمدُ دائمًا على مفهوم النمذجة البعدية. وتُعرَف النمذجة البعدية كنظامٍ قائمٍ على الاسـترجاع يدعمُ وصول طلبات البحث ذات الحجم الكبير. ويجبُ أن يتمَّ تصميمُ وتمثيل البيانات في مستودع ولكن أيضًا تعزِّز معالجة الاستفسارات المعقدة المتعددة الأبعاد. وفي كثيرٍ من الأحيان؛ فإنَّ مخطط النجمة ومخطط snowflake هما الوسـيلة التي يتمُّ بها تنفيذُ النمذجة البعدية في مستودعات البيانات.

يُعَدُّ مخططُ النجوم (ويُشـار إليه أحيانًا باسـم مخطط الانضمام إلى النجوم) أبسطَ وأكثرَ أنماط نمذجة الأبعاد اسـتخدامًا. ويحتوي مخطط النجوم على جدول حقائق مركزي مُحاط بعِدَّة جداول أبعاد ومتصلٌ بها (Adamson, 2009). ويحتوي جدول الحقائق على عددٍ كبيرٍ من الصفوف التي تتوافـقُ مع الحقائـق المرئية والروابط الخارجية (أيْ: مفاتيح خارجيـة). ويحتوي جدول الحقائق

على السِّمات الوصفية اللازمة لإجراء تحليل القرار وإعداد تقارير الاستعلام، ويتمُّ استخدام المفاتيح الخارجيــة للارتباط بجداول الأبعاد. وتتكوَّن ســمات تحليل القرار مــن مقاييس الأداء، والمقاييس التشغيلية، والمقاييس المجمَّعة (مثل: حجم المبيعات، ومُعدَّلات الاحتفاظ بالعملاء، وهوامش الرِّبح، وتكاليــف الإنتاج، ومعدَّل الخردة) وجميع المقاييس الأخرى اللازمة لتحليل أداء المنظمة. وبعبارة أخرى؛ فإنَّ جدول الحقائق يتناول بشكلٍ أساسيٍّ ما يدعم مستودع البيانات لتحليل القرارات.

إنَّ جداول الحقائــق المركزية المحيطة (والمرتبطة عبر المفاتيح الخارجية) هي جداول الأبعاد. وتحتــوي جداول الأبعــاد على تصنيف وتجميع المعلومات حول صفــوف الحقائق المركزية. كما تحتوي على السِّــمات التي تَصِفُ البيانات الموجودة داخل جدول الحقائق، وتتناول كيفية تحليل البيانات وتلخيصها. وإضافةً إلى ذلك؛ فإنها تحتوي أيضًا على علاقة واحد - إلى - مُتعدِّد في جدول الحقائق المركزية، وعند اســتخدام الاستعلام، يتمُّ استخدامُ الأبعاد لوضع القيم العددية في شرائح مــن ١-٦ في جدول الحقائــق لتلبية الاحتياجات المتعَلِّقة بالمعلومــات. وقد تمَّ تصميمُ المخطط النجمي؛ بهدف توفير وقت الاستعلام والحصول على استجابة سريعة؛ إضافةً إلى بساطته وسهولة صيانة هياكل قاعدة بيانات القراءة فقط. ويوضِّح الشــكل ٣a-١٠مخططاً نجمياً بســيطاً، ويُعَدّ المخطط النجمي حالةً خاصةً لمخطط مجموعة الثلج (Snowflake).

وبالنسبة لمخطط مجموعة الثلج (Snowflake)؛ فهو عبارة عن ترتيب منطقي للجداول في قاعدة بيانات مُتعدِّدة الأبعاد؛ بحيث يشــبه المخطط الارتباطي مجموعة الثلج في الشــكل، ويرتبط مخطط مجموعــة الثلــج (Snowflake) ارتباطًا وثيقًــا بالمخطط النجمي عن طريق جــداول حقائق مركزية (تكــون واحدة فقــط عادةً) والتي ترتبط بأبعــاد متعددة. ومع ذلك ففي مخطــط مجموعة الثلج (Snowflake) يتمُّ تســوية الأبعاد إلى عِدة جداول مرتبطة ببعضها، في حين يتمُّ تقوية أبعاد مخطط النجمة، مع تمثيل كلِّ بُعْد بجدول واحد، ويظهر مخطط مجموعة ثلج بسيط في الشكل ٣b-١٠.



شكل ٣-١٠: (a) مخطط النجمة، و(b) مخطط كتلة الثلج

رؤية فنية ٣-١: مستودعات البيانات المُستضَافة

يحتوي مستودع البيانات المستضاف على نفس الوظيفة، إن لم يكن أكثر؛ فهو يعمل كمستودع بيانات في الموقع، غير أنه لا يستهلك موارد الحاسب في أماكن عمل العميل. ويُقدِّم مستودع البيانات المستضاف مزايا ذكاء الأعمال مطروحًا منها تكلفة تحديثات الحاسب، وتحديثات الشبكة، وتراخيص البرامج والتطوير الداخلي والدعم الداخلي والصيانة.

ويُقدِّم مستودع البيانات المستضاف المزايا التالية:

- يتطلَّب الحَدَّ الأدنى من الاستثمار في البنية التحتية.
- يُحرِّر القدرة على الأنظمة الداخلية.
- يُحرِّر التدفق النقدي.
- يُقدِّم حلولاً قويةً بأسعار معقولة.
- يُمكِّن الحلول القوية التي تساعد على النمو.
- يُقدِّم مُعَدَّات وبرامج ذات جودة أفضل.
- يوفِّر اتصالات أسرع.
- يُمكِّن المستخدمين من الوصول إلى البيانات من المواقع البعيدة.
- يسمح للشركة بالتركيز على الأعمال الأساسية.
- يفي باحتياجات التخزين لأحجام كبيرة من البيانات.

وعلى الرغم من مُميِّزات مستودع البيانات المستضاف؛ فإنَّ ذلك لا يستلزم أن يكون مناسبًا لكلِّ منظمة. فبعض الشركات الكبيرة التي تتجاوز عائداتها ٥٠٠ مليون دولار قد تخسر الأموال إذا كان لديها بالفعل بنيةٌ تحتيةٌ داخلية وموظفون في مجال تقنية المعلومات. وعلاوةً على ذلك؛ فإن الشركات التي ترى أن التحوُّل النموذجي يحدث عن طريق الاستعانة بمصادر خارجية في تطبيقات فقدان السيطرة على التحكُّم في البيانات الخاصة بهم لا تفضل استخدام مُقدِّم خدمة ذكاء الأعمال. وأخيرًا؛ فإن أكثر الحجج شيوعًا والتي تعيق تطبيق مستودع البيانات المستضافة، هو أنه قد يكون من غير الحكمة الاستعانة بمصادر خارجية للتطبيقات الحساسة لأسباب تتعلق بالأمان والخصوصية.

تحليل البيانات في مستودع البيانات:

بمجرد تخزين البيانات بشكلٍ مناسبٍ في مستودعات البيانات؛ يمكِن استخدامها بطرق مختلفة لدعم القرارات التنظيمية. ويُمكِن القول بأن تقنية OLAP هي تقنية تحليل البيانات

الأكثر اســتخدامًا في مســتودعات البيانات، وقــد تزايد الإقبال على هذه التقنية؛ بســبب الزيادة الهائلــة في حجم البيانات والاعتراف بالأعمال القيِّمة للتحليلات المبنية على البيانات. وببســاطة؛ فإن OLAP هو نهجٌ للإجابة بسرعة عن الأسئلة المخصَّصة عن طريق تنفيذ استفسارات تحليلية متعدِّدة الأبعاد ضد مستودعات البيانات التنظيمية (أي: مستودعات البيانات، وDMs).

OLAP مقابل OLTP:

يُســتَخدَم مصطلــح OLTP (وهو نظام معالجة المعاملات عبر الإنترنــت) في نظام المعاملات المسئول بشكلٍ أساسيٍّ عن التقاط وتخزين البيانات المتعلِّقة بوظائف العمل اليومية، مثل: ERP، وCRM، وSCM، وPOS، وغيرها. ويتعامل نظام OLTP مع احتياجات الأعمال الحساسة، وأداء معاملات الأعمال اليومية بطريقة أوتوماتيكية، وتشــغيل التقارير فوريًّا، والتحليل الروتيني. غير أنَّ هذه الأنظمة ليســت مُصمَّمة للتحليل المُخصَّص والاستفسارات المعقدة التي تتعامل مع عددٍ من عناصر البيانات. ومن ناحيةٍ أخرى؛ فقد تمَّ تصميمُ OLAP لمعالجة هذه الحاجة؛ من خلال توفير تحليلٍ مُخصَّصٍ للبيانات التنظيمية بشــكلٍ أكثر فعاليةً وكفاءةً. ويعتمد كلٌّ من OLAP و OLTP بشكلٍ كبيرٍ على بعضهما؛ إذ يستخدم OLAP البيانات التي تمَّ التقاطها بواسطة OLTP، ويقــوم OLTP آليًّا بإجــراء عمليات الأعمال التي تتمُّ إدارتها بواســطة القــرارات التي يدعمها OLAP. ويقدِّم الجدول ٣-٥ مقارنةً متعددة المعايير بين OLTP وOLAP.

جدول ٣-٥: مقارنة بين OLTP وOLAP

المعايير	OLTP	OLAP
الغرض	للقيام بمهامِّ العمل اليومية	لدَعْــم صُنْــع القــرار وتقديــم إجابات لاستفسارات الأعمال والإدارة
مصدر البيانات	قاعــدة بيانــات المعاملات (مســتودع بيانات عادي يركِّز بشــكلٍ أســاسي على الكفاءة والاتساق)	مســتودع البيانات أو DM (مســتودع بيانــات غير طبيعي يركِّز في المقام الأول على الدقة والاكتمال)
التقارير	تقارير دورية، وروتينية ومركزة بدقة	تقارير واستعلامات مُخصَّصة ومتعددة الأبعاد ومركزة على نطاق واسع
متطلبات المصدر	قواعد البيانات المترابطة (العلائقية) العادية	المعالجات، وقواعد البيانات المتخصِّصة ذات السّعة الكبيرة

المعايير	OLTP	OLAP
سرعة التنفيذ	سريع (تسجيل المعاملات التجارية والتقارير الروتينية)	بطيء (موارد مكثفة، ومعقدة، واستعلامات واسعة النطاق)

عمليات OLAP:

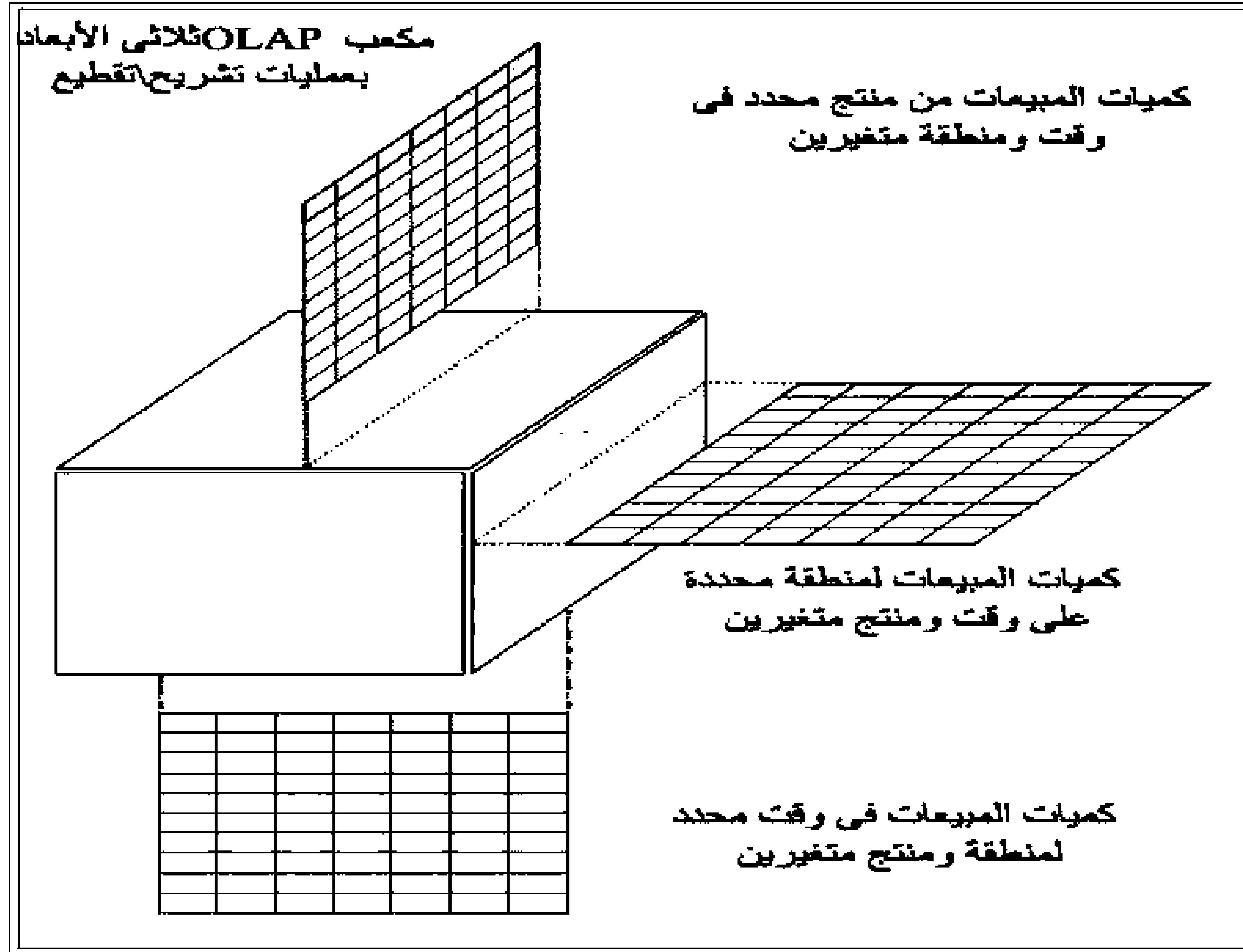
يعتمد الهيكلُ التشغيلي الرئيسي في OLAP على مفهومٍ يُسمَّى المكعب. والمكعب في OLAP؛ هو عبارة عن بنية بيانات متعددة الأبعاد (فعلية أو افتراضية) والتي تسمح بتحليلٍ سريعٍ للبيانات، كما يُمكِن تعريف المكعب بأنه القدرة على معالجة البيانات وتحليلها بكفاءة من وجهات نظر مُتعدِّدة، ويهدف ترتيب البيانات في مكعبات إلى التغلُّب على القصور في قواعد البيانات المترابطة (العلائقية)؛ إذ إنَّ قواعد البيانات (المترابطة) العلائقية لا تناسب التحليل الفوري لكميات كبيرة من البيانات بشكلٍ جيد؛ بل إنها بدلًا من ذلك تتيحُ الفرصة للتلاعب بالسِّجلات (من خلال إضافة وحذف وتحديث البيانات) والتي تمثل سلسلةً من المعاملات. وعلى الرغم من وجود العديد من أدوات كتابة التقارير لقواعد البيانات العلائقية؛ فإنَّ هذه الأدوات تكون بطيئة عند تنفيذ استعلام متعدد الأبعاد يشمل العديد من جداول قاعدة البيانات.

وباستخدام OLAP؛ يُمكِن للمحلِّل التنقل عبر قاعدة البيانات والشاشة لمجموعةٍ فرعية مُعيَّنة من البيانات (وتطويرها بمرور الوقت) عن طريق تغيير اتجاهات البيانات وتعريف العمليات الحسابية. وتُسمَّى هذه الأنواع من التنقلات التي يقوم بها المستخدِم من خلال تحديد شرائح (عن طريق التناوب) والتنقل لأسفل/ لأعلى (عبر التجميع والتصنيف) أحيانًا بـ «شريحة ونرد»، وتتضمَّن عمليات OLAP الشائعة الاستخدام شريحةً ونرداً، والتنقيب لأسفل، ولفة المتابعة والمحور.

- **الشريحة:** هي عبارة عن مجموعةٍ فرعيةٍ من صفيف مُتعدِّد الأبعاد (وعادةً تمثيل ثنائي الأبعاد)، هذه المجموعة تطابق قيمةً واحدةً لبُعدٍ واحدٍ (أو أكثر) من الأبعاد غير الموجودة في المجموعة الفرعية. ويبَيِّن الشكل ٣-١١ عملية تشريح بسيطة على مكعب ثلاثي الأبعاد.
- **حجر النَّرد:** النَّرد هي عبارة عن شريحة على أكثر من بُعدَين من مكعب البيانات.
- **التنقل لأسفل/ لأعلى:** التنقل لأسفل أو لأعلى هو تقنية OLAP مُحدَّدة؛ إذ يتنقل المستخدِم بين مستويات البيانات التي تتراوح من الأكثر إيجازًا (أعلى) إلى الأكثر تفصيلًا (أسفل).
- **لفة المتابعة:** يتضمَّن اللف إلى الأعلى حساب جميع علاقات البيانات لواحد أو أكثر من الأبعاد.

وللقيام بذلك؛ قد يتمُّ تعريف علاقة حسابية أو صيغة.

- **المحور:** ويُستَخدَم لتغيير اتجاه الأبعاد لتقرير أو عرض صفحة الاستعلام المخصَّص.



شكل ٣-١١: عمليات تقطيع على مكعب بيانات بسيط ثلاثي الأبعاد

أسئلة مراجعة على القسم ٣-٦:

١- اذكر فوائد مستودعات البيانات.

٢- اذكر عِدَّة معايير لاختيار بائع مستودع البيانات، واذكر سبب أهميتها.

٣- ما هو OLAP؟ وكيف يختلف عن OLTP؟

٤- ما هو المكعب؟ وماذا يعني التنقل لأسفل، واللف، والشريحة، والنَّرد؟

٣-٧ قضايا تنفيذ عملية مستودعات البيانات:

إنَّ عملية مســتودعات البيانات بشــكلٍ عام تتطلَّب جهدًا هائلًا يجب التخطيط له وتنفيذه وفقاً للأساليب المتَّبعة. غير أنَّ دورة حياة المشروع لها عِدَّة جوانب ولا يمكن لأيِّ شخصٍ أن يكون

خبــيراً في كلِّ المجالات. ولذلك فنحن هنا بصَدَد مناقشــة أفــكار وقضايا مُحدَّدة لها صِلة بعملية مستودعات البيانات.

يرغــبُ الناس في معرفة مدى نجاح مبادرات ذكاء الأعمال ومخازن البيانات الخاصة بهم مقارنةً بمبادرات الشركات الأخرى. وقد اقترح كلٌّ من Ariyachandra وWatson (٢٠٠٦أ) بعض معايير نجاح ذكاء الأعمال ونشاط مستودعات البيانات. كما قام كلٌّ من Watson، وGerard، وGonzalez، و Haywood، و(1999) Fenton بالبحث في حالات فشل مستودع البيانات. وقد أظهرت نتائجهم أنَّ الناس يعرفون «الفشــل» بطرق مختلفة، وهذا ما أكَّده كلٌّ من Ariyachandra، و(Watson 2006أ). وقد قام معهدُ مســتودعات البيانات (tdwi.org) بتطوير نموذجٍ متكاملٍ لمســتودعات البيانات تستطيع المؤسســة تطبيقه لقياس مدى تطوُّرها. ويقدِّم النموذج وسيلةً سريعةً لقياس مدى صلاحية مكان مستودعات البيانات الخاصة بالمنظمة في الوقت الحالي، وكذلك الأماكن التي تحتاجُ إلى الانتقال إليها فيما بعد. ويتكوَّن نموذجُ النضج من سِتِّ مراحل، وهي: مرحلة ما قبل الولادة، الرضاعة، الطفولة، المراهقة، البلوغ، الحكمة.

وترتفع قيمةُ الأعمال مع تقدُّم مستودع البيانات خلال كلِّ مرحلة من مراحل النجاح. ويتمُّ تحديدُ المراحل من خلال عددٍ من الخصائص، بما في ذلك النطاق، والهيكل التحليلي، والتصوُّرات التنفيذية، وأنواع التحليــلات، والإشراف، والتمويــل، وبرنامج التقنية، وإدارة التغيــير، والإدارة، ولمزيدٍ من التفاصيل انظر: Eckerson، وHackathorn، وMcGivern، وTwogood، و(Watson(2009، و(Eckerson (2003.

تتسمُ مشاريعُ مستودع البيانات بالعديدِ من المخاطر. كما أنَّ معظم هذه المخاطر موجودةٌ أيضًا في مشــاريع تقنية المعلومات الأخرى، ولكن مخاطر مســتودعات البيانات أكثر خطورةً؛ لأن مســتودعات البيانات غالية الثمن وتتطلب الكثير من الوقت والموارد، ومشــاريع واسعة النطاق. لذلــك؛ فإن التطوير الناجح لمســتودع البيانات يتطلَّب تقييــم كلِّ خطر في بداية المشروع. ومن المهمِّ التفكير بعناية في مختلف المخاطر، وتجنُّب المشكلات التالية:

- **الخطأ في اختيار الرُّعاة:** فأنت بحاجةٍ إلى راعٍ تنفيذيٍّ له نفوذٌ على الموارد الضرورية لدعم واستثمار مخزن البيانات. كما أنك تحتاجُ أيضًا إلى قائدٍ تنفيذيٍّ للمشروع، ولا بُدَّ أن يكون هذا القائد شخصًا اكتســب احترام المسؤولين التنفيذيين الآخرين، ولديه شكوكٌ مقبولةٌ حول التقنية، كما أنه لا بُدَّ أن يتصف بالحَزْم مع المرونة. وإضافةً إلى ذلك؛ فإنك أيضًا تحتاج إلى مدير IS/IT ليرأس المشروع.
- **تحديد توقعات يَصعُب الوفاء بها:** أنت لا تريد إحباط المديرين التنفيذيين بإخبارهم بحقائق معينة. كلُّ مشروع من مشاريع مستودعات البيانات يتكوَّن من مرحلتين:

- **المرحلة الأولى:** هي مرحلة البيع؛ إذ تقوم بتسـويق المشروع داخليًّا عن طريق بيع الفوائد لِمَنْ يستطيعون الوصول إلى الموارد المطلوبة.
- **المرحلـة الثانيـة:** وتتمثل في الاجتهاد؛ من أجل الوفاء بتوقعـات المرحلة الأولى، فبمجرد أن يكون لديك ما بين ١ و٧ ملايين دولار؛ فإننا نأمل أن تتمكَّن من تحقيق ذلك.
- **الانخراط في اتباع نهجٍ سُـلوكي غير جيد:** ببسـاطة لا تقل بأن مسـتودعات البيانات ستسـاعد المديريـن في اتخاذ قرارات أفضل؛ لأن ذلك ضمنياً يعني أنَّ قراراتهم السَّـابقة غير جيدة حتى الآن. ولكـن قـم بالترويج لفكرة أنهم سـيكونون قادرين على الحصول عـلى المعلومات التي يحتاجونها للمساعدة في صُنْع القرار.
- **تحميل المُسـتودَع بالمعلومات لمجرد توفُّرها:** لا تجعل مسـتودع البيانـات مُكِبًّا للبيانات؛ لأن هذا من شأنه أن يُبطئ استخدام النظام بغير داعٍ. وهناك اتجاهٌ نحو الإحصاء والتحليل بشكلٍ فوري. ويجب إغلاق مستودعات البيانات في الوقت المناسب.
- **الاعتقـاد أن تصميمَ قاعدة بيانات؛ من أجل مسـتودعات البيانات هـو نفس تصميم قاعدة بيانـات المعاملات:** وبشـكلٍ عام؛ فإنه ليس كذلـك. فالهدف من مسـتودعات البيانات هو الوصول إلى المجاميع بدلًا من السـجلات الفردية أو القليلـة، كما هو الحال في أنظمة معالجة المعامـلات. ويختلف المحتـوى أيضًا كما هو واضـح في كيفية تنظيم البيانـات؛ إذ تميل نُظُم إدارة قواعـد البيانات (DBMS) إلى أن تكون غير مفرطـة، وطبيعية، ومترابطة، في حين تكون مستودعات البيانات زائدة عن الحاجة، وليست طبيعية، ومُتعدِّدة الأبعاد.
- **اختيار مدير مسـتودع البيانات، والذي يكون مُوجِّهًا للتقنية بدلًا من المسـتخدِم:** يتمثل أحد مفاتيح نجاح مسـتودع البيانات في فهم ضرورة حصول المسـتخدِمين على احتياجاتهم، وليس التقنية المتقدِّمة من أجل التقنية.
- **التركيـزُ على البيانـات الداخلية التقليدية الموجَّهة للسـجل، وتجاهل قيمـة البيانات الخارجية والنصوص والصور، وربما الصوت والفيديو:** تأتي البيانات بأشكال عديدة، ويجب أن تكون متاحةً للأشخاص المناسبين في الوقت المناسب وبالشكل الصحيح. ويجبُ أن يتمَّ فهرستها بشكلٍ صحيح.
- **تقديـم البيانات باسـتخدام تعريفات متداخلة ومذهلة:** يُعَدُّ تطهـيرُ البيانات جانبًا مهمًّا من مسـتودعات البيانات. ويشـمل التوفيق بين تعريفات وتنسـيقات البيانـات المتعارضة على مستوى المنظمة، غير أنَّ ذلك قد يكون صعبًا من الناحية السياسية؛ نظرًا لأنه ينطوي عادةً على تغيير المستوى التنفيذي.

- **الإيمانُ بالقدرة على الوفاء بوعود الأداء، والقدرة والقابلية للتطوير:** تتطلَّب مســتودعات البيانات بشكلٍ عام سعةً وسرعةً أكبر مما هو مرصود في الميزانية. أيْ: التخطيط للمستقبل على نطاق واسع.
- **الاعتقاد أنَّ المشــكلات قد انتهت عندما يكون مســتودع البيانات جاهزًا:** إذ تحتاج مشــاريع BI/ DSS إلى التطوير بشــكلٍ مســتمر. فكلُّ نَشْر هو تكرار لعملية النماذج الأوَّلية. وســيكون هناك حاجةٌ دائمةٌ لإضافة المزيد من مجموعات البيانات المختلفة إلى مستودع البيانات؛ إضافةً إلى أدوات تحليلية إضافية للمجموعات الموجودة والإضافية من صُنَّاع القرار، كما يجب تخطيط الطاقة العالية والميزانيات السنوية؛ لأن النجاح يُولِّد نجاحاً، فبناء مستودعات البيانات عملية مستمرة.
- **التركيــز على التنقيــب في البيانات المخصَّصة، وتقديم التقارير الدورية بدلًا من التنبيهات:** إن التطوُّر الطبيعي للمعلومات في مســتودع البيانات هــو: (١) التنقيب في البيانات من الأنظمة القديمة، وتطهيرها، وتغذية المستودع بها، (٢) دَعْم التقارير المخصَّصة حتى يمكِنك التعرُّف على مــا يريده النــاس، (٣) تحويل التقارير المخصَّصة إلى تقارير مجدولــة بانتظام. وهذه العملية المتمثلة في معرفة ما يريده الناس من أجل توفيره تبدو طبيعية؛ لكنها ليســت الأمثل أو حتى عملية. فالمديرون مشــغولون وهــم في حاجة إلى الوقت لقراءة التقاريــر؛ ولذلك فإن أنظمة التنبيه أفضل من أنظمة التقارير الدورية، ويُمكِنها أن تجعل مهمة مســتودع البيانات حاسمةً. وتقوم أنظمة التنبيه بمراقبة تدفق البيانات إلى المســتودع وإبلاغ جميع الأشــخاص الرئيســيين الذين يحتاجون إلى معرفة أيِّ حَدَث مهم فور وقوعه.

في العديدِ من المؤسســات، لن يكون مســتودع البيانــات ناجحًا إلا إذا كان هنــاك دَعْمٌ قويٌّ مــن الإدارة العليــا من أجــل تطويره، وإذا كان هنــاك بطلٌ للمشروع ارتقــى إلى أعلى في الهيكل التنظيمي. فعلى الرغم من أنَّ هذا الأمر من المحتمل أن يكون صحيحًا في أيِّ مشروعٍ كبير لتقنية المعلومات؛ فإنه مهمٌّ بشــكلٍ خاصٍّ لتنفيذ مســتودع البيانات. إذ يؤدي التنفيذ الناجح لمستودع البيانات إلى إنشــاء هيكل إطار عمل قد يســمح بتحليل القرارات في جميع أنحاء المؤسســة، وفي بعــض الحــالات يوفر أيضًا SCM شــاملاً؛ وذلك من خــلال منح الوصــول إلى المعلومات الخاصة بعملاء وموردِّي المنظمة. وقد سهَّل تنفيذُ مستودعات البيانات المستندة إلى الويب (والتي تُسمَّى أحيانًــا Webhousing) الوصول إلى كميات هائلة من البيانات، ولكن من الصعب تحديد الفوائد الثابتة المرتبطة بمســتودع البيانات. إذ يتمُّ تعريــف الفوائد الثابتة على أنها منافع للمنظمة التي يُمكِــن التعبيرُ عنها من الناحية النقدية. ويُوجَد لدى العديد من المنظمات موارد محدودة لتقنية المعلومات غير أن الأولوية يجب أن تُعطى للمشــاريع. وقد يســاعد الدعم الإداري والبطل القوي للمشروع في ضمان أن يتلقى مشروع مستودع البيانات الموارد اللازمة للتنفيذ الناجح. كما أن موارد

مستودع البيانات قد تكون مكلفةً بشكلٍ كبير، وفي بعض الحالات تتطلب معالجات عالية الجودة وزيادات كبيرة في أجهزة تخزين الوصول المباشر. وقد تحتوي مستودعات البيانات المستندة إلى الويب أيضًا على متطلبات أمان خاصة؛ لضمان وصول المستخدمين المصرَّح لهم فقط إلى البيانات.

تُعَدُّ مشاركة المستخدمين في تطوير البيانات ونماذج الوصول عاملًا حاسمًا في نجاح تطوير مستودع البيانات؛ إذ إنه في أثناء نمذجة البيانات يكون عامل الخبرة مهمًّا لتحديد البيانات المطلوبة، وتحديد قواعد العمل المرتبطة بالبيانات؛ إضافةً إلى تحديد المجموعات والحسابات الأخرى التي قد تكون ضروريةً. وهناك حاجةٌ لنمذجة الوصول لتحديد كيفية استرجاع البيانات من مستودع البيانات، كما أنها تساعدُ في التعريف المادي للمستودع من خلال المساعدة في تحديد البيانات التي تتطلب فهرسة. وقد تشير أيضًا إلى ما إذا كانت DMs التابعة مطلوبةً لتسهيل استرجاع المعلومات. وإضافةً إلى ذلك؛ فإن هناك حاجة إلى مهارات فريق لتطوير وتنفيذ مستودع البيانات، بما في ذلك المعرفة المتعمِّقة لتقنية قواعد البيانات وأدوات التطوير المُستَخدَمة. وتشيرُ أنظمة المصدر وتقنية التطوير - كما سبق أن ذكرنا - إلى العديد من المدخلات والعمليات المُستَخدَمة لتنزيل وصيانة مستودع البيانات. وتمثل الحالة العملية ٣-٤ مثالًا ممتازًا على تطبيق واسع النطاق لمستودع بيانات متكامل من قِبل حكومة الولاية.

حالة عملية ٣-٤

EDW يساعد على رَبْط وكالات الدولة في ولاية ميشغان

تؤثر إدارة Michigan للتقنية والإدارة والميزانية (DTMB) على كلِّ مجالٍ من مجالات الحكومة؛ من خلال خدمة العملاء، وتحسين الموارد، والاستخدام المبتكر للمعلومات والتقنية، ويعتمد ما يقرب من ١٠٠٠٠ مستخدم في ٥ أقسام رئيسة و٢٠ وكالة وأكثر من ١٠٠ مكتب على EDW للقيام بمهامِّهم بشكلٍ أكثر فاعلية وخدمة أفضل للمقيمين في Michigan. ويحقق EDW مليون دولار في يوم عمل في الفوائد المالية.

وقد ساعد EDW ميشيغان في تحقيق فوائد مالية سنوية بقيمة ٢٠٠ مليون دولار ضِمن قسم صحة المجتمع وَحْده؛ إضافةً إلى ٧٥ مليون دولار أخرى في السنة ضمن دائرة الخدمات الإنسانية (DHS). وتشمل هذه المدخرات فوائد سلامة البرنامج، وتفادي التكاليف؛ بسبب تحسُّن النتائج، وتجنب العقوبات، والكفاءة التشغيلية، واسترداد المدفوعات غير الملائمة ضمن برنامج Medicaid الخاص بها. ويوفِّر مستودع البيانات DHS في Michigan (DW) معلومات فريدة ومبتكرة حاسمة لعمل الوكالة بكفاءة

من المســتوى الإستراتيجي والتكتيكي. فعلى مدى السنوات العشر الماضية حققت شركة DW نسـبة فعالية التكلفة ١٥:١، وتُسـهِم المعلومات الموحَّدة من DW الآن في جميع وظائف DHS تقريبًا، بما في ذلك التسليم الدقيق والمحاسبة للفوائد المقدَّمة ممَّا يقرب من ٢,٥ مليون عميل من موظفي الدَّعم العام.

وكانت Michigan طموحةً في محاولاتها لحَلِّ مشـكلات الحياة الحقيقية من خلال المشـاركة المبتكـرة والتحليلات الشـاملة للبيانات؛ إذ كان نهجها تجـاه BI/DW دائمًا «مشروعًـا» (على مسـتوى الولاية) بطبيعته، بدلًا من وجـود مِنصَّات منفصلة من BI/DW لكلِّ منطقة عمل أو وكالة حكومية. وقد استفادت Michigan من كميات هائلة من البيانات لإنشـاء أساليب مبتكرة لاستخدام BI/DW، وتقديم حلول مؤسسية تتسمُ بالكفـاءة والمِصدَاقية باسـتخدام قنوات متعددة؛ وذلك من خـلال إزالة الحواجز التي تحول دون مشاركة بيانات المؤسسات عبر وحدات الأعمال.

أسئلة للمناقشة:

١- لماذا تستثمر الدولة في بنية أساسية كبيرة ومُكلِّفة لتقنية المعلومات (مثل EDW)؟

٢- ما هو حجم وتعقيد EDW المُستَخدَمة من قِبل وكالات الدولة في ولاية Michigan؟

٣- ما هي التحَدِّيات والحَلّ المقترح والنتائج التي تمَّ الحصول عليها من EDW؟

Sources: Compiled from TDWI Best Practices Awards 2012 Winner, Enterprise Data Warehousing, Government and Non-Profit Category. Michigan Departments of Technology, Management & Budget (DTMB), Community Health (DCH), and Human Services (DHS). TDWI What Works, 34, 22; michigan.michigan.gov.

مستودعات البيانات الضخمة وقابلية التطوير:

يحتاج مستودعُ البيانات إلى دَعْم قابلية التوسُّع بالإضافة إلى المرونة، وتتمثل القضايا الرئيسية المتعلقـة بقابلية التطويـر فيما يلي: مقدار البيانـات الموجودة في المسـتودع، ومدى سرعة نموِّ مستودع التخزين، وعدد المستخدِمين المتزامنين، وتعقيد طلبات المستخدِمين، كما يجب أن يتسع مستودعُ البيانات أفقيًّا وعموديًّا. وسوف ينمو المستودعُ كدالة لنمو البيانات والحاجة إلى توسيع المستودع لدعم وظائف الأعمال الجديدة. وقد يكون نموُّ البيانات ناتجًا عن إضافة بيانات الدورة الحالية (على سبيل المثال: نتائج هذا الشهر) أو البيانات السابقة.

وقــد وَصَــف هيكس (٢٠٠١) قواعد البيانــات الضخمة ومخازن البيانــات. وتقوم Walmart باســتمرار بزيادة حجم مستودع البيانات الضخم الخاص بها. ومن المعتقد أن Walmart تستخدم مستودعًا يحتوي على مئات من terabytes من البيانات لدراسة اتجاهات المبيعات، وتتبُّع المخزون، وتنفيــذ مهــامّ أخرى. وقــد قامت IBM مؤخرًا بنــشر مؤشرها الخاص بالمســتودع الذي يبلغ ٥٠ تيرابايــت (IBM، 2009). الجديرُ بالذكر أن وزارة الدفاع الأمريكية تســتخدم مســتودعًا للبيانات يبلغ ٥ بيتابايت لحفظ الســجلات الطبية لــعدد ٩ ملايين من الأفراد العســكريين. كما أن CNN أيضًا لديها مستودع بيانات بحجم Petabyte؛ وذلك نظرًا للتخزين المطلوب لأرشفة لقطات الأخبار.

وبالنظر إلى أنَّ حجم مستودعات البيانات يتوسَّع بمعدل أُسِّي؛ فإن القابلية للتطوير هي قضية مهمــة. ويُقصَد بقابلية التوسُّــع الجيد أن الاســتعلامات ووظائف الوصــول إلى البيانات الأخرى ستنمو (بشــكلٍ مثالي) خطيًّا مع حجم المستودع. انظر: Rosenberg (2006) للتعرُّف على طرق تحســين أداء الاستعلام. وفي الممارسة العملية، تمَّ تطويرُ أســاليب مُتخصِّصة لإنشاء مستودعات بيانات قابلة للتطوير، غير أنَّ قابلية التوسُّــع صعبة عند إدارة مئات terabytes أو أكثر. فبيانات بحجــم terabytes تحتــوي على قدرٍ كبيرٍ من القصور الذاتي، وتحتلُّ الكثير من المســاحة المادية، كما أنها تتطلَّب أجهزة حاسب قوية. وتستخدم بعض الشركات المعالجة المتوازية لإدارة بياناتهم، في حين يســتخدم البعضُ الآخر الفهرســة الذكية وخطط البحث. كما يقوم البعض بنشر بياناتهم عــبر مخازن البيانات الفعليــة المختلفة. ومع اقتراب المزيدِ من مســتودعات البيانات من حجم petabyte يستمرُّ تطوير حلول أفضل وأحسن لقابلية التوسع.

وقــد عالج Hall (2002) أيضًا قضايــا قابلية التطوير، وتُعَدُّ شركة AT&T شركةً رائدةً في مجال نشر واستخدام مستودعات البيانات الضخمة، وقد استطاعت AT&T اكتشاف الاحتيالية لاستخدام بطاقــات الاتصال والتحقيق في المكالمات المتعلقة بعمليات الخطف والجرائم الأخرى؛ وذلك بفضل مســتودع البيانات الذي تبلغ ســعته ٢٦ تيرابايت. كما أنها اســتخدمت القدرة على حساب ملايين أصــوات الاتصال من مشــاهدي التلفزيون الذين يختارون بطل برنامــج American Idol المُقبِل. وللحصول على عيِّنة من تطبيقات مســتودعات البيانات الناجحة، انظر: Edwards (2003). وقد درس Jukic و(2004) Lang الاتجاهات والقضايا المُحدَّدة المتعلقة باستخدام الموارد الخارجية في تطوير ودَعْم مســتودعات البيانات وتطبيقات ذكاء الأعمال، كما أشــار Davison (2003) إلى أن الاستعانة بمصادر خارجية في مجال تقنية المعلومات كانت تتزايد بمعدل ٢٠ إلى ٢٥٪ سنويًّا، وعند النظر في نقل مشــاريع مســتودعات البيانات إلى الخارج؛ يجب النظــر بعناية إلى الثقافة والأمن (ولمزيدٍ من التفاصيل انظر: Jukic & Lang, 2004).

أسئلة مراجعة على القسم ٣-٧:

١- ما هي مهامُّ تنفيذ DW الرئيسية التي يمكن تنفيذها بالتوازي؟

٢- اذكر مع الشرح المبادئ التوجيهية الأكثر دقة في تنفيذ DW.

٣- ما هي أهمُّ المخاطر والقضايا التي يجب مراعاتها، وربما تجنبها عند تطوير مستودع بيانات ناجح؟

٤- ما هي القابلية للتطوير؟ وكيف تطبَّق على DW؟

٣-٨ إدارة مستودع البيانات، والمشكلات الأمنية، والاتجاهات المستقبلية:

توفِّر مستودعات البيانات مَيْزةً تنافسيةً متميزةً للمؤسسات التي تُنشِئها وتستخدمها بفاعلية. ونظرًا لحجمها الكبير وطبيعتها الجوهرية؛ فإنَّ مستودع البيانات يتطلب مراقبةً قويةً بشكلٍ خاصٍّ؛ للحفاظ على كلٍّ من الكفاءة والإنتاجية المرضيتين. وتنطوي الإدارة الناجحة وإدارة مستودع البيانات على المهارات والكفاءة التي تتجاوز ما هو مطلوب من مسؤول قاعدة البيانات التقليدي (DBA). ويجب أن يكون مسؤول مستودع البيانات (DWA) متوافقاً مع تقنيات البرامج العالية الأداء والأجهزة والشبكات. كما يجب أيضًا أن يمتلك رؤية أعمال رائعة. ولما كانت مستودعات البيانات تغذِّي أنظمة ذكاء الأعمال وDSS والتي تساعد المديرين في اتخاذ قرارتهم، وَجَبَ على مسؤول مستودع البيانات أن يكون على درايةٍ بعمليات اتخاذ القرار لتصميم هيكل مستودع البيانات وصيانته بشكلٍ مناسب. ومن المهمِّ بشكل خاص أن يحافظ DWA على المتطلبات والإمكانات الحالية لمستودع البيانات؛ في حين يوفر في الوقت نفسه مرونةً للتحسينات السريعة. وأخيرًا، يجب أن تمتلك DWA مهارات اتصال ممتازة. وللحصول على وَصْف للاختلافات الأساسية بين DBA وDWA، انظر: Benander، وBenander، وFadlalla، وJames (2000).

يُعَدُّ أمان المعلومات وخصوصيتها من الاهتمامات الرئيسة والمهمة بالنسبة لمُستخدِمي مستودع البيانات. وقد أصدرت الحكومة الأمريكية لوائح (على سبيل المثال: قواعد الخصوصية والضمان لـ Gramm-Leach-Bliley، والتأمين الصحي وقانون المساءلة لعام ١٩٩٦ [HIPAA])، ووضعت متطلباتٍ إلزاميةً في إدارة معلومات العملاء. وبالتالي؛ يجب على الشركات إنشاء إجراءات أمنية فعَّالة؛ ولكنها مرنةٌ في الوقت ذاته لتتوافق مع العديد من أنظمة الخصوصية. ووفقًا لما ذكره كلٌّ من Elson و(2005) LeClerc؛ فإنه يجب أن يركِّز الأمان الفعَّال في مستودع البيانات على أربعةِ مجالات رئيسية:

١- وَضْع سياسات وإجراءات أمنية فعَّالة للشركات. ويجب أن تبدأ السياسة الأمنية الفعَّالة من الإدارة التنفيذية العليا، وينبغي إبلاغُ جميع الأفراد داخل المؤسسة.

٢- تنفيذ إجراءات وتقنيات أمنية منطقية لتقييد الوصول. وتتضمَّن هذه الإجراءات التحقق من المستخدِم، وعناصر التحكم في الوصول، وتقنية التشفير.

٣- تقييد حرية الوصول إلى محيط مركز البيانات.

٤- إنشاء عملية مراجعة فعَّالة للرَّقابة الداخلية مع التركيز على الأمن والخصوصية.

وسوف يتمُّ تحديدُ تطورات مستودعات البيانات على المدى القريب؛ وذلك من خلال عوامل ملحوظة (مثل: حجم البيانات، وزيادة درجة التأهب لوقت الاستجابة، وتنوع وتعقيد أنواع البيانات) وعوامل أخرى أقل وضوحًا (مثل: متطلبات المستخدِم النهائي التي لم تستوفِ البيانات، وبطاقات الأداء المتوازن، وإدارة البيانات الرئيسية، وجودة المعلومات). وفي ضوءِ هذه العوامل المُحَرِّكة اقترح كلٌّ من Moseley (2009)، و Agosta (2006)أن تعتمد اتجاهات مستودعات البيانات على البساطة، والقيمة، والأداء.

مستقبل مستودعات البيانات:

لقد كان مجالُ مستودعات البيانات مجالًا حيويًّا في مجال تقنية المعلومات في العقدين الأخيرين، وتبرهن الأدلة في عالم BI/BA والبيانات الضخمة أن أهمية المجال سوف تزداد فقط حتى تصبح أكثر إثارةً للاهتمام. وفيما يلي بعض المفاهيم والتقنيات التي ظهرت مؤخرًا والتي ستلعب دورًا مهمًّا في تحديد مستقبل مستودعات البيانات.

الاستعانة بمصادر مختلفة (آليات للحصول على البيانات من مصادر متنوعة ومتفرقة):

- **شبكة الإنترنت، ووسائل التواصل الاجتماعي، والبيانات الضخمة:** إن الطفرة الأخيرة في استخدام شبكة الإنترنت لأغراض شخصية إضافةً إلى الأعمال التجارية، إلى جانب الاهتمام الكبير بوسائل التواصل الاجتماعي يوفر فرصًا للمُحلِّلين للاستفادة من مصادر بيانات غنية جدًّا. وبسبب الحجم الهائل والسرعة وتنوُّع البيانات؛ فقد ظهر مصطلحٌ جديدٌ لتسمية هذه الظاهرة، وهو مصطلح: «البيانات الضخمة». وللاستفادة من البيانات الضخمة؛ فإن الأمر يتطلب تطوير تقنيات BI/BA الجديدة والمُحسَّنة؛ مما سيؤدي إلى ثورة في عالم مستودعات البيانات.

Sources: Compiled from Thornton, M., & Lampa, M. (2002). Hosted data warehouse. Journal of Data Warehousing, 7(2), 27-34; Thornton, M. (2002, March 18). What about security? The most common, but unwarranted, objection to hosted data warehouses. DM Review, 12(3), 30-43.

- **برمجيــات مفتوحة المصدر:** يتزايدُ اســتخدام أدوات البرمجيات مفتوحة المصدر إلى مســتوى غير مســبوق في التخزين، وذكاء الأعمال، وتكامل البيانات. وهناك أسـباب وجيهة للزيادة في البرمجيات المفتوحة المصدر المُستَخدَمة في مستودعات البيانات (Russom, 2009)، وهي:
  ١- أدَّى الركود إلى زيادة الاهتمام بالبرامج المفتوحة المصدر المنخفضة التكلفة.
  ٢- وصلت الأدوات مفتوحة المصدر إلى مستوى أعلى في النضج.
  ٣- تقوم البرمجيات مفتوحة المصدر بتقوية برامج المشاريع التقليدية دون استبدالها.
- **SaaS (كخدمــة برمجيــة)، «نمــوذج ASP الموسَّــع»:** تُعَــدُّ SaaS طريقــةً مبتكــرةً لنــشر تطبيقــات أنظمــة المعلومــات؛ إذ يقــوم مُوفِّــر الخدمــة بمَنْح العمــلاء ترخيصاً لاســتخدام تطبيقاتــه عنــد الطلب لاســتخدام تطبيقاته كخدمة عنــد الطلب (وعادةً مــا يكون ذلك عبر الإنترنــت). ويُمكِــن لمُــورِّدي برامج SaaS اســتضافة التطبيق عــلى الخــوادم الخاصة بهم أو تنزيــل التطبيق إلى موقع المســتهلك. والمضمون أن SaaS هي النســخة الجديدة والمُحسَّــنة مــن نموذج ASP. وبالنســبة لعملاء مســتودع البيانات؛ فــإن العثور عــلى تطبيقات وموارد البرمجيــات القائمــة عــلى SaaS، والتي تلبِّــي الاحتياجــات والمتطلبات الخاصــة من الممكن أن يكــون أمــرًا صعبًا. ونظــرًا لأن عروض هذه البرامــج أصبحت أكثر مرونةً؛ فإن الاســتعانة بـ SaaS والاستخدام الفعلي لها كخيار لِمنصَّة مستودعات البيانات سوف يزداد أيضًا.
- **حوســبة ســحابية:** ربما تكون الحوسبة السَّــحابية هي الخيارُ الأحدث والأكثر إبداعًا في النظام الأســاسي الذي سيأتي على مدار ســنوات؛ إذ يتمُّ تجميعُ العديد من موارد الأجهزة والبرمجيات الافتراضيــة؛ بحيث يمكن تخصيصها بحرية للتطبيقات ومِنصات البرامج عند الحاجة إلى الموارد. مما يتيح لتطبيقات أنظمة المعلومات أن ترتفع ديناميكيًّا مع زيادة أعباء العمل. وعلى الرغم من أن تقنيات الحوســبة الســحابية والتقنيات الظاهرية المشــابهة معروفة جيدًا للتطبيقات التشــغيلية اليوم؛ فإنها بدأت الآن في استخدامها كبرامج مستودع بيانات مختارة. فالتخصيص الفعَّال للسَّــحابة يكون مفيداً بشــكلٍ خاص عندما يختلف حجم مستودع البيانات بشكلٍ غير متوقع؛ مما يجعل تخطيط السَّعة أمرًا صعبًا.
- **بحيرات البيانات:** مع ظهور البيانات الضخمة ظهر برنامج بيانات جديد يُسمَّى: بحيرة البيانات، وهــو عبارة عن موقع تخزين كبير يمكنه الاحتفاظ بكميات هائلة من البيانات (والتي غالبًا ما تكون غير مهيكلة) في نَسَــقها الأصلي/ الخام لاستهلاك التحليلات المستقبلية/ المحتملة. فبينما يقوم مســتودعُ البيانات بمســتودعات البيانات المنظمة؛ فإن بحيرة البيانات تخزِّن جميع أنواع

البيانات. في حين أن كليهما عبارة عن آليات لمستودعات البيانات؛ فمستودع البيانات يتمحور حول البيانات المنظمة/ المجدولة، أما بحيرة البيانات تناسب جميع أنواع البيانات. وعلى الرغم مــن كثرة ما قِيل وكُتِب عن العلاقة بين الاثنين (بعض هذه الكتابات، وبعض ما قيل يوحي بأن بحيرة البيانات، هي الاســم المستقبلي لمســتودعات البيانات)؛ فإنَّ بحيرة البيانات ليست بديلًا لمستودع البيانات؛ بل يُكمِّل كلٌّ منهما الآخر. وفيما يلي الفكرة التطبيقية ٣-٢ والتي نركِّز من خلالها على شرح بحيرات البيانات ودورها في مجال مستودعات البيانات والتحليلات التجارية.

رؤية فنية ٣-٢: بحيرات البيانات:

مــع ظهور ظاهــرة البيانات الكبيرة؛ ظهر مصطلــحٌ جديدٌ وهو «بحــيرة البيانات». ويعتقد الكثــيرون أن بحــيرة البيانات هي نفســها مســتودع البيانــات القديم؛ ولكن بــشرط أن يكون جيــدًا. ويشــير الافــتراض الأســاسي إلى أن الطريقــة القديمة لمســتودعات البيانات لا تناســب (أولا تكفي) في عصر البيانات الكبيرة، وبالتالي؛ فإن هناك حاجةً إلى طريقة جديدة لتخزين/ إدارة البيانــات؛ مما مهَّد الطريق لبحــيرات البيانات. وعلى الرغم من أن معظمهم يعتقدون أن بحيرة البيانات هي الطريقة المُثلى، وأنها جاهزة لعمل قفزة كبيرة في مجال مســتودعات البيانات؛ فإنَّ البعض الآخر له تحفُّظات على استخدامها (وربما شكَّك في جدواها)، واصفاً إياها بأنها مستنقع.

إذًا، ما هي حقيقة بحيرة البيانات؟

إنَّ بحيرة البيانات ببســاطة هي موقعُ تخزين كبيرٍ يمكنه الاحتفاظ بكميات هائلة من البيانات (مهيكلة أو غير مهيكلة أو شــبه مهيكلة) بتنسيقها الأصلي/ الخام للاستخدام المستقبلي المحتمل. في حين أن مســتودع البيانات لا يُخزِّن إلا البيانات المنظمة في الجداول أو الملفات أو المجلدات، كما أن بحــيرة البيانات تســتخدم بنيةً غير مُحدَّدة المعالم (أيْ: غير مهيكلــة) لتخزين جميع أنواع البيانات. وتتمثل القواسم المشتركة الرئيسية بين بحيرة البيانات ومستودع البيانات في أنَّ كلًا منهما يُعَدُّ آليات لمستودعات البيانات، وعلى العكس؛ فإن الاختلاف الرئيسي هو أن مستودع البيانات يتعَلَّق بالبيانات المنظمة / المُجدولة، في حين أن بحيرة البيانات تتعلق بجميع أنواع البيانات (مثل: البيانات الكبيرة).

عــلى الرَّغم من أن تعريــف بحيرة البيانات يتغيَّر بــين علماء البيانات؛ فــإنَّ التعريف الأكثر اســتخدامًا هو تعريف James Dixon، مؤســس شركة Pentaho وCTO، والذي ارتبط اســمه بمصطلــح بحيرة البيانــات. وفيما يلي الطريقة التي يَصِف بهــا (Dixon, 2010) بحيرة البيانات؛ إذ يقول: إذا كنتَ تفكر في ســوق البيانات كمخــزن للمياه المعبأة - من حيث تنظيفها وتعبئتها

وتنظيمها لسـهولة الاسـتهلاك - فإنَّ بحيرة البيانات عبارة عن مجموعة كبيرة من المياه في حالة طبيعية، وتدفق إليها المحتويات من أحد المصادر لملء البحيرة، وفي هذه الحالة يمكِن لمستخدمي البحيرة المتنوعين أن يفحصوها، أو يغوصوا فيها، أو حتى يأخذوا عينات منها.

ولعلَّ أفضل طريقة لتوصيف بحيرة البيانات؛ هي مقارنتها بمستودع بيانات في جدول مُتعدِّد الأبعاد، وفيما يلي الجدول (٣-٦) وهو عبارة عن جدول ملخص (يتبعه وصفٌ مُوجَز) للأبعاد الأكثر شيوعًا لعمل مقارنة بين بحيرة البيانات ومستودع البيانات (Dull, 2016; Campbell, 2015).

جدول ٣-٦: مقارنة بسيطة بين مستودع البيانات وبحيرة البيانات

البُعد	مستودع البيانات	بحيرة البيانات
طبيعة البيانات	منظمة ومعالجة	أي بيانات أصلية / أولية
المعالجة	مُخطط على الكتابة (SQL)	تصميم
سرعة الاسترجاع	سريع جداً	المخطط على القراءة (NoSQL)
حجم التكلفة	مكلفة للبيانات الكبيرة	بطيئة مُصمَّمة للتكلفة المنخفضة
المرونة	أقل مرونةً، والتكوين ثابت	التخزين مَرِن للغاية، والتكوين مَرِن
الحداثة/ الخبرة	ليس جديداً/ ناضج	جديد جداً/ ناضج
الأمان	مُأمَّن بطريقة جيدة	لم يُأمَّن جيداً بعد
المستخدمون	مُحترفو الأعمال	علماء البيانات

طبيعة البيانات: يقوم مسـتودع البيانات بمستودعات البيانات التي تمَّ تصميمها/ تجميعها/ هيكلتهـا فقـط؛ في حين أنَّ بحـيرة البيانات تقوم بتخزين كافـة أنواع البيانات المهيكلة، وشِـبه المهيكلة؛ إضافةً إلى غير المهيكلة بتنسيقها الأصلي/ الأوَّلي.

المعالجـة: قبل تنزيل البيانات في مسـتودع البيانات؛ نحتاجُ أولًا إلى إعطائها شـكلًا وهيكلًا، أيْ: إننا نحتاجُ إلى وَضْع نموذجٍ لها في المخطط النَّجمي أو مخطط مجموعة الثلج، وهو ما يُسمَّى مخطط الكتابة. أمَّا مع بحيرة البيانات؛ فإننا نقوم فقط بتنزيل البيانات الأولية، كما هي، وعندما نكون مسـتعدين لاسـتخدام البيانات؛ فإننا نعطيها شكلًا أو تصميمًا، وهو ما يُطلَق عليه مخطط القراءة. وهما طريقتان مختلفتان للغاية في المعالجة.

سرعــة الاســترجاع: لأكثر من عِقدين وهناك تطويرٌ مســتمرٌ للعديد من القواعد الحســابية لتحسين سرعة استرجاع البيانات من مستودعات البيانات التي تتميَّز بأنها مليئة وكبيرة. وتشمل هذه التقنيات المشغلات، والتمثيل العمودي للبيانات، ومعالجة قواعد البيانات. وحتى الآن؛ فإنَّ عملية استرجاع البيانات (التي يُمكِن أن تكون بأيِّ شكلٍ أو أسلوب، بما في ذلك النصُّ غير المنظم) تُعَدُّ نشاطًا يتطلب وقتًا كبيرًا.

تكلفة التخزين: تُعَدُّ إحدى المَيْزات الرئيســة لتقنيات البيانات الكبيرة، مثل Hadoop هي أنَّ تكلفة مستودعات البيانات منخفضة نسبيًّا مقارنةً بمستودع البيانات، وهناك سببان رئيسان لذلك:

- **السبب الأول:** Hadoop وهو برنامجٌ مفتوح المصدر؛ ولذلك فإن الترخيص ودَعْم المجتمع مجاني.
- **السبب الثاني:** أنه تمَّ تصميمُ Hadoop؛ بحيث يتمُّ تثبيته على أجهزة السِّلع المنخفضة التكلفة.

المرونة: بالتعريف يُعَدُّ مســتودع البيانات مستودعًا منظمًا للغاية. وليس من الصعب تقنيًّا تغيير بنيته؛ ولكن قد يستغرق هذا الأمرُ وقتًا طويلًا؛ نظرًا لجميع العمليات التجارية المرتبطة به. ومن ناحيةٍ أخرى؛ فإنَّ بحيرة البيانات تفتقرُ إلى بنية مستودع البيانات؛ مما يعطي علماء البيانات والمُطوِّرين القدرة على تهيئة نماذجهم وطلباتهم وتطبيقاتهم بسرعة وإعادة تشكيلها.

الحداثــة/ الخبرة: كانــت التقنيات المتوفرة في مســتودع البيانات موجودةً منــذ فترة طويلة. ومعظــم الابتــكارات التي حدثت تمَّ إنجازها في العشرين إلى الثلاثــين عامًا الأخيرة. ولذلك فإنه لا يُوجَد تحديث في مستودعات البيانات، وإن وُجِدَ فهو قليلٌ جدًّا (مع استبعاد التقنية اللازمة لتعبئة واستخدام «البيانات الكبيرة» داخل مستودع البيانات). وعلى الجانب الآخر؛ فإنَّ بحيرات البيانات جديدة، وتتقدم من خلال الابتكار والحداثة لتصبح هي التقنية السائدة لمستودعات البيانات.

الأمان: نظراً لوجود تقنيات مستودع البيانات منذ عقود طويلة؛ فإنَّ القدرة على تأمين البيانات في مســتودع البيانات أكــبر بكثيرٍ منها في بحيرة البيانات. ومع ذلك؛ فإنه تجدر الإشــارة إلى الجهد الكبير الذي يُبذل الآن في مجال الأمان في صناعة البيانات الضخمة. ولا يتعلق الأمرُ بما إذا كان أمن بحيرات البيانات سوف يُلبِّي متطلبات محترفي التحليلات وغيرهم من المستخدمين النهائيين، أم لا.

المســتخدمون: لقد كان شــعار العالَــم التحليلي لفــترة طويلة؛ هو «تحليــل وفهم الأعمال للجميع». وقد أنشأنا مستودع البيانات ووجَّهنا الدعوة للجميع لاستخدامه؛ ولكن هل استخدموه بالفعل؟ والإجابة هي أن نسبة مَنْ استخدموه لا تتعَدَّى ٢٠-٢٥٪ فقط. هل بحيرة البيانات هي نفس صيحة مستودع البيانات؟ هل سنقوم بإنشاء بحيرة البيانات وندعو الجميع إلى استخدامها؟

ربما يحدث ذلك في المستقبل. أمَّا في الوقت الحالي؛ فإن بحيرة البيانات بوضعها الحالي، ومرحلة النضج التي وصلت إليها تُعَدُّ هي الأنسب لعلماء البيانات.

وباختصارٍ؛ فإن بحيرة البيانات تختلف عن مستودع البيانات. فبحيرة البيانات ليست هي مستودع البيانات (كما اقترح البعض)، كما أنها ليست بديلًا له. فكلاهما مطلوبان ولذلك؛ فقد تمَّ تحسينهما من أجل وسائط البيانات المختلفة والمهام/ الأغراض المتنوعة. ولذلك؛ فإن كلا الاثنين بحاجةٍ إلى التعايش معًا في عالم التحليلات (على الأقل لفترة من الوقت حتى تصل بحيرات البيانات إلى مستوى من النضج يُحقِّق ما وصلت إليه مستودعات البيانات من الجودة في الوقت الحاضر). والهدف هو تصميم واستخدام كلٍّ منهما بشكلٍ صحيحٍ للمهامِّ التي يؤديها؛ بحيث يتمُّ استخدام أفضل خيار للعمل، والذي قد يكون مزيجٌ من مستودعات البيانات وبحيرة البيانات.

البُنية التحتية (التصميم - الأجهزة والبرمجيات - التحسينات):

- **العمودية (طريقة جديدة لمستودعات البيانات والوصول إليها في قاعدة البيانات):** إنَّ نظام إدارة قواعد البيانات باستخدام الأعمدة الموجَّهة (والمعروف أيضًا باسم قاعدة البيانات العمودية)، هو نظامٌ يقوم بتخزين جداول البيانات كأجزاء من أعمدة البيانات بدلًا من أنها صفوف من البيانات (وهي الطريقة التي يستخدمها معظم RDBMS). بمعنى أن تخزين قواعد البيانات يتمُّ بشكلٍ عموديٍّ؛ فهذه البيانات تُخزَّن بواسطة أعمدة بدلًا من الصفوف (حيث يتمُّ تخزينُ كافة قِيَم عمود واحد بشكلٍ متتابع على ذاكرة القرص). ومثل هذا الهيكل يعطي نوعاً من التحكُّم أكثر دقة في RDBMS. فهو يستطيعُ الوصول إلى الأعمدة المطلوب الاستعلام عنها فقط دون أن يكون هناك اضطرارٌ إلى الوصول إلى كلِّ أعمدة الصف؛ إذ يعمل بشكلٍ أفضل بكثير فيما يخصُّ الاستعلامات التي تحتاج إلى نسبةٍ صغيرة من الأعمدة في الجداول الموجودة بها؛ ولكنه يؤدي عمله بشكلٍ أسوأ عندما تكون هناك حاجةٌ إلى أعمدة كثيرة؛ بسبب الجهد الزائد المطلوب لربط كافة الأعمدة معًا لتشكيل مجموعات النتائج. وعادةً ما تتعلق المقارنات بين مخططات بيانات الأعمدة الموجهة والصفوف الموجهة، بكفاءة الوصول إلى القرص الثابت؛ للقيام بتنزيل عمل مُحدَّد (والذي يُعَدُّ واحدًا من أكثر العمليات التي تستغرق وقتًا طويلًا في الحاسب). وبناءً على المهمة المطروحة، يمكن لأحدهما أن يتمتع بمزايا أكثر من الآخر. ويكون نظام الأعمدة الموجهة أكثر كفاءة عندما:

١- يجب حساب الإجمالي من خلال العديد من الصفوف؛ ولكن فقط لمجموعة فرعية صغيرة؛ بحيث تكون هذه المجموعة أقل من إجمالي أعمدة البيانات؛ إذ إنَّ قراءة هذه المجموعة الفرعية الأصغر من البيانات تكون أسرع من قراءة كلِّ البيانات.

٢- يتمُّ تزويدُ جميع الصفوف بقِيَم جديدة من أحد الأعمدة في وقتٍ واحد؛ نظرًا لأن بيانات العمود يمكِن كتابتها بكفاءة واستبدال بيانات الأعمدة القديمة دون لمس أي أعمدة أخرى للصفوف.

في حين يكون نظامُ الصفوف الموجّهة أكثر كفاءةً عندما:

١- تكون أعمدة كثيرة من صف واحد مطلوبة في نفس الوقت، وعندما يكون حجم الصف صغيراً نسبيًّا؛ بحيث يمكن استرداد الصف بأكمله باستخدام قرص واحد.

٢- يتمُّ كتابة صَفٍّ جديدٍ إذا تمَّ توفير كافة بيانات العمود في نفس الوقت؛ إذ يمكِن كتابة الصف بأكمله باستخدام قرصٍ واحد.

وإضافــةً إلى ذلك؛ فإنه نظــرًا لأن البيانات المخزّنة في أحد الأعمدة تكون من نوعٍ مُوحَّد؛ فإنَّ ذلك يفســح المجال بشكلٍ أفضل للضغط، بمعنى أنه يتوفر مساحة كبيرة لتخزين بيانات الأعمدة الموجّهة، وهذه المساحة لا تتوفر في بيانات الصفوف الموجهة، ومثل هذا الضغط المُحْكَم للبيانات مــن شــأنه أنه يقلِّل من حجــم التخزين؛ مما يجعله أكــثر البدائل التي تجد قبــولًا من الناحية الاقتصادية لمتابعة التخزين في الذاكرة أو الحالة الصلبة.

- **مستودعات البيانات الفورية:** تنطوي عملية مستودعات البيانات الفورية (RDW) على دورة تجديد مســتودع البيانات الموجود؛ وذلك من أجل تحديث البيانات الأكثر تكرارًا (ويكون ذلك تقريبًــا في نفس الوقت الذي تصبح فيه البيانات متاحةً في قواعد البيانات التشــغيلية). فهذه الأنظمة (أنظمة RDW) يمكنها الوصول إلى تحديث البيانات بشكل فوري تقريبًا؛ إذ يستغرق وقت استجابة البيانات عادةً من دقائق إلى ساعات. ومع انخفاض وقت الاستجابة؛ فإن تكلفة تحديث البيانات تزداد بشــكلٍ كبير. وهناك حاجة إلى التقدُّم مســتقبلًا في العديد من المظاهر التقنية (بدءًا من اكتســاب البيانات تلقائيًّا وصولًا إلى أدوات البرمجيات الذكية) لجعل RDW حقيقة واقعة بتكلفة معقولة.

- **أجهــزة مســتودعات البيانــات (حلول الــكلِّ في واحد لــدى DW):** يتكوَّن جهاز مســتودع البيانات من مجموعةٍ متكاملةٍ من الخوادم، ووحدات التخزين، ونظام (أو أنظمة) التشــغيل، وأنظمة إدارة قواعد البيانات، والبرمجيات المثبتة بشــكلٍ مُسبَق، والمُثبَّتة من أجل مستودعات البيانات. ومن الناحية العملية؛ فإن أجهزة مســتودع البيانات توفر حلولًا من أجل مســتودع البيانات المتوســطة لدى ســوق مســتودع البيانات الكبيرة؛ إذ تقدِّم أداءً منخفضَ التكلفة على أحجــام البيانات في نطاقٍ من تيرابايت إلى نطاق البيتابايت. ويســتخدم معظم مورِّدي أجهزة مستودعات البيانات تصميمات معالجة متوازية كثيفة من أجل تحسين الأداء. وعلى الرغم من

أن معظم مُورِّدي قواعد البيانات ومستودعات البيانات يوفرون الأجهزة في الوقت الحاضر، غير أنَّ الكثيرين يعتقدون أن تيراداتا كانت أول من قدَّم منتجًا يتمثل في جهازٍ لمستودعات البيانات التجارية. وأكثر ما يُلاحَظ حاليًا هو ظهورُ حُزَم مستودعات البيانات؛ إذ يقوم الموزِّدون بدَمْج برمجياتهم وقواعد بياناتهم كمِنصَّة لمستودعات البيانات.

أمَّا من ناحية مزايا أجهزة مستودع البيانات؛ فإنَّ التكلفة الإجمالية للحصول عليها وتملُّكها منخفضةٌ للغاية، والتي تشمل تكاليف الشراء الأولية، وتكاليف الصيانة المستمرة، وتكلفة تغيير السّعة مع نمو البيانات. كما تُشكِّل تكلفة الموارد اللازمة لمراقبة وضبط مستودع البيانات نسبةً كبيرةً من إجمالي تكلفة المِلكية، وغالبًا ما تصل هذه النسبة إلى ٨٠٪. وإضافةً إلى ما سبق؛ فإن أجهزة DW تعمل على الحَدِّ من إدارة العمليات اليومية والمتمثلة في التشغيل والإعداد والتكامل. ونظرًا لأن أجهزة مستودعات البيانات توفر حلَّ المورد الواحد؛ فإنها تتجه إلى تحسين الأجهزة والبرامج بشكلٍ أفضل داخل الجهاز؛ إذ يعمل هذا التكامل الموحَّد على زيادة فرص الاندماج الناجح واختبار نظم إدارة قواعد البيانات (DBMS) والتخزين ونظام التشغيل عن طريق تجنُّب بعض مشكلات التوافق التي تنشأ عن الحلول التي يقدِّمها مُورِّدون مُتعدِّدون. كما يوفر جهاز مستودع البيانات نقطة اتصال واحدة لحَلِّ المشكلة ومسار ترقية أبسط بكثيرٍ لكلٍّ من البرامج والأجهزة.

- **تقنيات وممارسات إدارة البيانات:** تتضمَّن بعض الاحتياجات المُلِحَّة للجيل القادم من برنامج مستودع البيانات التقنيات والممارسات التي لا نفكِّر فيها عمومًا على اعتبار أنها جزءٌ من النظام الأساسي؛ إذ يحتاجُ العديدُ من المستخدمين على وجه الخصوص إلى تحديث أدوات إدارة البيانات التي تعمل على معالجة البيانات؛ لتصبح صالحةً للاستخدام من خلال مستودع البيانات. ويحمل المستقبل نموًّا قويًّا لإدارة البيانات الرئيسية (MDM). ورغم كون هذا المفهوم جديداً نسبيًّا غير أنه بالغ الأهمية، وسوف يَحْظى بشعبية كبيرة لعِدَّة أسباب، وتشمل ما يلي:

١- إن الدَّمج المُحكَم مع الأنظمة التشغيلية يحتاجُ إلى إدارة بيانات رئيسية (MDM).

٢- إن معظمَ مستودعات البيانات لا تزال تفتقرُ إلى وظائف إدارة البيانات الرئيسية وجودة البيانات.

٣- يجبُ أن تكون التقاريرُ التنظيمية والمالية دقيقةً ونظيفةً تمامًا.

- **تقنية معالجة قاعدة البيانات (وَضْع نظام الحلول الحسابية حيث تكون البيانات):** تشيرُ معالجة قواعد البيانات (وتُسمَّى أيضًا تحليلات قواعد البيانات) إلى مدى التكامل الحسابي لتحليلات البيانات في مستودع البيانات. وبمجرد القيام بذلك؛ فإن البيانات والتحليلات تتكاملُ لتصبح ضِمن بيئة واحدة. فوجود كلا الاثنين في بيئةٍ واحدة يُزيد من كفاءة إجراء التحليل

الحسابي المكثف. واليوم تعتمدُ العديدُ من أنظمة اتخاذ القرارات على قواعد البيانات الكبيرة، مثل تلك المُستَخدَمة في الكشف عن الاحتيال على بطاقات الائتمان وإدارة مخاطر البنوك الاستثمارية؛ إذ تُستَخدَم هذه التقنية نظرًا لإظهارها تَحسُّنًا كبيرًا في الأداء يتفوق على الطرق التقليدية في بيئة اتخاذ القرار، والتي يُشكِّل فيها الوقت عاملًا مهمًّا. وتُعَدُّ معالجة قواعد البيانات عمليةً معقدةً إذا ما قورنت بالطريقة التقليدية لإجراء التحليلات؛ إذ يتمُّ نقل البيانات خارج قاعدة البيانات (وغالبًا ما تكون في شكل ملفات مُسَطَّحَة تتألف من صفوف وأعمدة) إلى بيئة تحليلية منفصلة (مثل SAS Enterprise Modeler، أو Statistica Data Miner، أو IBM SPSS Modeler) للمعالجة.

وتُعَدُّ المعالجة داخل قواعد البيانات أكثرَ منطقيةً بالنسبة لبيئات التطبيقات العالية الإنتاجية الفورية، والتي تتضمَّن كَشْف الاحتيال، وسجل الائتمان، وإدارة المخاطر، ومعالجة المعاملات، والتسعير وتحليل الهامش، والتجزئة الدقيقة المستندة إلى الاستخدام، واستهداف الإعلانات السلوكية، ومحركات التوصية، مثل تلك التي تستخدمها مؤسسات خدمة العملاء لتحديد الإجراءات التالية الأفضل، ويتم تنفيذ ومعالجة البيانات داخل قاعدة البيانات على اعتبار أنها مَيْزة، من قِبل العديد من مُوَرِّدي مستودعات البيانات الرئيسيين، بما في ذلك تيراداتا (دَمْج قدرات SAS analytics في أجهزة مستودعات البيانات)، وIBM Netezza، وEMC Greenplum، وSybase، من بين البرامج الأخرى.

- **تقنية التخزين في الذاكرة (نقل البيانات في الذاكرة؛ حتى تتمَّ المعالجة بشكلٍ أسرع):** تستخدم أنظمة قواعد البيانات التقليدية مثل RDBMS، محركات الأقراص الصلبة الفعلية لمستودعات البيانات بشكلٍ مثالي لفترة طويلة من الوقت؛ إذ يقوم نظام إدارة قاعدة البيانات بتنزيل البيانات (أو أجزاء من البيانات) إلى الذاكرة الرئيسية عند طلب عملية متعلقة بالبيانات بواسطة أحد التطبيقات، ويعالجها، ويستجيب مرةً أخرى للتطبيق. وعلى الرغم من أن البيانات (أو أجزاء من البيانات) تُخزَّن مؤقتًا في الذاكرة الرئيسية في نظام إدارة قاعدة البيانات؛ فإنَّ موقع التخزين الأساسي يظل عبارة عن قرص مغناطيسي ثابت. وفي المقابل؛ فإن نظام قاعدة البيانات في الذاكرة يقوم بالاحتفاظ بالبيانات بشكلٍ دائمٍ في الذاكرة الرئيسية. وعندما تُطلب عملية تتعلق بالبيانات بواسطة تطبيقٍ ما؛ فإن نظام إدارة قواعد البيانات يصل مباشرةً إلى البيانات الموجودة بالفعل في الذاكرة الرئيسية، ويعالجها، ويستجيب مرةً أخرى للتطبيق الذي يطلبه. إنَّ هذا الوصول المباشر إلى البيانات في الذاكرة الرئيسة يجعل معالجة أوامر البيانات أسرع بكثير من الطريقة التقليدية.

لذا؛ فإنَّ الفائدة الرئيسة للتقنية في الذاكرة (بل قد تكون الفائدة الوحيدة لها) تتمثل السرعة المذهلة التي تصل بها إلى البيانات. أمَّا العيوب فتشمل: تكلفة الحصول على ذاكرة رئيسة كبيرة للغاية (فحتى لو تمَّ الحصول عليها بتكلفة أقل، إلا أنه لا يزال هناك الكثيرُ لتوفير ذاكرة رئيسية كافية لاحتواء جميع بيانات الشركة)، إضافةً إلى الحاجة إلى إستراتيجيات معقدة لاستعادةً البيانات (لأن الذاكرة الرئيسية مُتقلِّبة ويمكن محوها بطريق الخطأ).

- **نظم إدارة قواعد البيانات الجديدة:** تتكوَّن مِنصَّة مستودع البيانات من عِدَّة مكونات أساسية، أهمُّها نظام إدارة قواعد البيانات (DBMS). ويُعَدُّ ذلك أمرًا طبيعيًّا فقط؛ لأن نظام إدارة قواعد البيانات (DBMS) هو مُكوِّن النظام الأساسي؛ إذ إنَّ معظم العمل يجب إنجازه بغرض تنفيذ نموذج بيانات وتحسينه لأداء الاستعلام. ولذلك؛ فإنَّ نظام إدارة قواعد البيانات (DBMS) هو المكان الذي يتوقع أن تحدث فيه العديدُ من ابتكارات الجيل التالي.
- **تحليلات مُتقدِّمة:** هناك طرق تحليلية مختلفة يمكن للمستخدِمين اختيارها في أثناء انتقالهم إلى ما وراء الأساليب الأساسية المستندة إلى OLAP والتحليلات المتقدِّمة. إذ يختار بعضُ المستخدمين طرقًا تحليليةً متقدمةً تعتمدُ على التنقيب في البيانات، والتحليلات التنبؤية، والإحصاءات، والذكاء الاصطناعي، وما إلى ذلك. ومع ذلك؛ فمن الواضح أن غالبية المستخدِمين يختارون الأساليب المستندة إلى SQL. وسواءً اعتمدوا على SQL أم لا؛ فيبدو أن التحليلات المتقدِّمة تُعَدُّ من بين أهمِّ الأمور المُبشِّرة بنجاح مستودعات البيانات في الجيل التالي.

ومن الواضح أن مستقبلَ مستودعات البيانات مليءٌ بالوعود والتحَدِّيات الكبيرة. وإذ إنَّ عالم الأعمال أصبح أكثر عالميةً وتعقيدًا؛ فإن الحاجة إلى أدوات ذكاء الأعمال وأدوات استرجاع البيانات سوف تصبح أكثر من ذي قبل. كما يبدو أن أدوات وتقنيات تقنية المعلومات سريعة التحسُّن تتحرك في الاتجاه الصحيح لتلبية احتياجات أنظمة ذكاء الأعمال المستقبلية.

أسئلة مراجعة على القسم ٣-٨:

١- ما الخطوات التي يُمكِن للمؤسسة اتخاذها لضمان أمان وسِرية بيانات العميل في مستودع البيانات الخاص بها؟

٢- ما هي المهارات التي يجبُ أن تتوفر في DWA؟ ولماذا؟

٣- ما هي التقنيات الحديثة التي قد تشكِّل مستقبل مستودعات البيانات؟ ولماذا؟

٣-٩ إدارة أداء الأعمال:

تنتهـي العديدُ من تطبيقات مسـتودع البيانـات بتطوير نظـام إدارة أداء الأعمال (BPM). ففـي الأعمال والأدبيـات التجارية؛ نجد أنَّ BPM لديها عددٌ من الأسـماء، بما في ذلك إدارة أداء الشركات (CPM)، وإدارة أداء المؤسسـات (EPM)، وإدارة المشاريع الإستراتيجية (SEM). وقد تمَّ إنشـاءُ CPM بواسطة شركة Garner لتحليل السوق (gartner.com). وأمَّا مصطلح EPM؛ فهـو مصطلحٌ مرتبطٌ بعروض Oracle (oracle.com) بنفس الاسـم. وأمَّا SEM؛ فهو المصطلح الـذي تسـتخدمه SAP (sap.com). وفي هذا الفصل؛ يُفضَّـل مصطلح BPM على المصطلحات الأخرى؛ لأنه الأقدم، والأكثر استخدامًا بشكلٍ عام، والوحيد الذي لا يرتبط ارتباطًا وثيقًا بموفر الحل الواحد. ويشير مصطلح إدارة أداء الأعمال (BPM) إلى طريقة سَيْر الأعمال والمنهجيات والمقاييس والتقنيات المُسـتَخدَمة من قِبل الشركات لقياس أداء الأعمال ومراقبتها وإدارتها. ويشـمل ثلاثة مكونات رئيسية (Colbert, 2009):

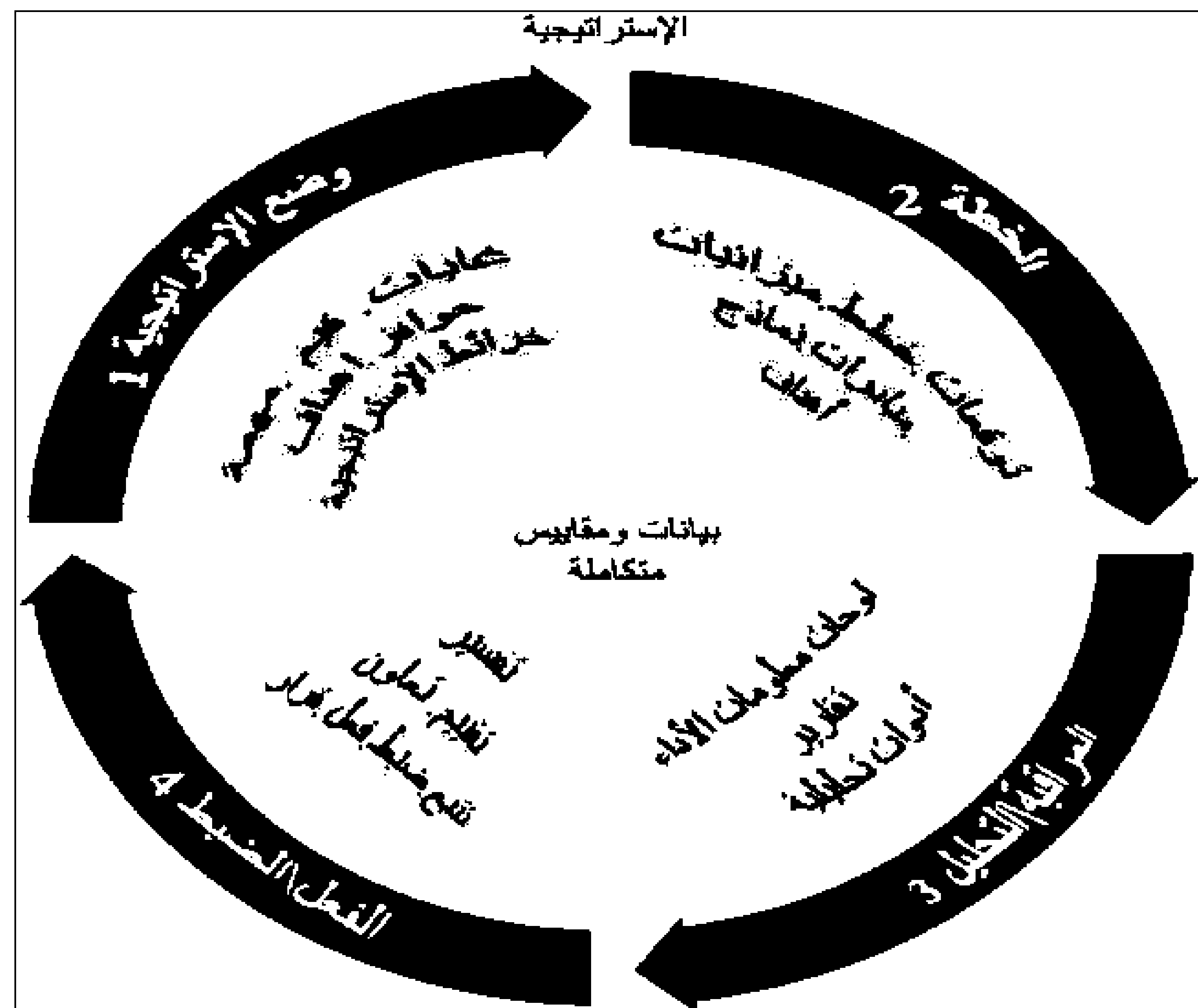
١- مجموعة مـن الإدارات المتكاملة في حلقة مغلقة، وعمليات التحليل (مدعومة بالتقنية) التي تتناول الأنشطة المالية، وكذلك الأنشطة التشغيلية.

٢- أدوات للشركات لتحديد الأهداف الإستراتيجية، ومن ثَمَّ قياسُ وإدارة الأداء في مقابل تلك الأهداف.

٣- مجموعة أساسية من العمليات، بما في ذلك التخطيط المالي والتشغيلي، والاندماج وتقديم التقارير، والنمذجة، والتحليل، ورَصْد مؤشرات الأداء الرئيسية (KPIs)، المرتبطة بالإستراتيجية التنظيمية.

دورة BPM المغلقة:

ربما يكون أهمُّ الاختلافات بين BPM وأيِّ أدوات وممارسات أخرى لذكاء الأعمال، هو التركيز عـلى الإسـتراتيجية الخاصـة به؛ إذ يشـمل BPM مجموعةً مـن العمليات المغلقـة التي تربط الإستراتيجية بالتنفيذ لتحسين أداء الأعمال (انظر: شكل ٣-١٢). ويُقصَد بالحلقة؛ أن الأداء الأمثل يتحقـق من خلال تحديد الأهداف والغايات (أيْ: وَضْع إسـتراتيجية)، وإنشـاء مبادرات وخطط لتحقيق تلك الأهداف (أيْ: خطة)، ومراقبة الأداء الفِعلي مقابل الأهداف والغايات (أيْ: المراقبة)، واتخاذ الفعل المناسب للإصلاح (أيْ: الفعل والضبط). فالطبيعة المستمرة والمتكرِّرة للدورة تعني أنَّ إكمال التكرار يؤدي إلى عملية جديدة ومُحسَّـنة (دعم الجهود المستمرة لتحسين العمليات). ونتناول في هذا القسم وَصْف هذه العمليات الأربع.



شكل ٣-١٢: حلقة مغلقة لدورة BPM

١- وَضْع الإستراتيجية: أين نريد أن نذهب؟

تُعرَّف الإســتراتيجية بشــكلٍ عام، بأنها خُطَّة عمل عالية المستوى، تشــمل فترةً زمنيةً طويلةً (غالبًا عدة ســنوات) لتحقيق هدفٍ مُحدَّد. وهي ضروريةٌ بشــكلٍ خاص في حالة وجود العديد من القيود (مدفوعة بأوضاع الســوق، وتوفُّر الموارد، والتعديلات القانونية/ السياســية) للتعامل معها في طريقها نحو تحقيق الهدف. في حين تُعرَّف الإســتراتيجية في إطار الأعمال التجارية بأنها فنٌّ وعلمٌ صياغة القرارات التي تساعد الشركات على تحقيق أهدافها. وبشكلٍ أكثر تحديدًا؛ فهي عمليــة تحديد وتوضيح مهمة المنظمة ورؤيتها وأهدافها ووَضْع خطط (على مســتويات مختلفة من الدقة - الإستراتيجية والتكتيكية والتشغيلية) لتحقيق هذه الأهداف.

وعادةً ما يتمُّ إنشــاءُ وتخطيط إســتراتيجيات الأعمال من قِبل فريقٍ من المسؤولين التنفيذيين في الــشركات (غالبًــا يقود هذا الفريق المديــرُ التنفيذي)، ويقوم مجلــس الإدارة بالموافقة عليها والتصريــح بها، ثم يقوم فريــق إدارة الشركة بتنفيذها تحت إشراف كبار المســؤولين التنفيذيين. وتوفر إستراتيجية الأعمال اتجاهًا عامًّا للمؤسسة وتُعَدُّ أول وأهمَّ عملية في منهجية BPM.

٢- الخطة: كيف نَصِل إلى هناك؟

عندمـا يعلمُ المديرون التنفيذيون ويفهمون ما (الأهداف، والأهداف التنظيمية)، سـيكونون قادريـن على التوصل إلى كيفية (وضع خطط تشـغيلية ومالية مُفصَّلـة)؛ بحيث تجيبُ الخطط التشـغيلية والمالية عن سؤالين: ما هي التكتيكات والمبادرات التي سيتمُّ اتباعها لتحقيق أهداف الأداء المُحدَّدة في الخطة الإستراتيجية؟ وما هي النتائج المالية المتوقعة لتنفيذ التكتيكات؟

تعملُ الخطة التشـغيلية عـلى ترجمة الأهداف والأهداف الإسـتراتيجية لمنظمةٍ ما إلى مجموعة مـن التكتيكات والمبادرات المُحـدَّدة جيدًا، ومتطلبات المـوارد، والنتائج المتوقعـة لبعض الوقت في المسـتقبل، وعادةً (ولكن ليس دائمًا) ما تكون هذه المدة عامًا. وتشـبه خطة التشـغيل في جوهرها خطة المـشروع التي تمَّ تصميمها؛ لضمان تحقيق إسـتراتيجية المؤسسـة، وتشـتمل معظم الخطط التشـغيلية على مجموعة من التكتيكات والمبادرات. وأمَّا مفتاح النجاح في التخطيط التشـغيلي فهو التكامل. وتقود الإستراتيجية إلى التكتيكات، وتقود التكتيكات إلى النتائج. وبشكلٍ أساسي يجبُ ربط التكتيكات والمبادرات المُحدَّدة في خطة تشغيلية بشكلٍ مباشر بالأهداف والغايات الرئيسية في الخطة الإسـتراتيجية. وإذا لم يكن هناك ارتباطٌ بين تكتيك فردي وهدف أو أهداف إستراتيجية أو أكثر؛ فإنه ينبغي على الإدارة أن تتساءل عمَّا إذا كان التكتيك والمبادرات المرتبطة به مطلوبةً فعلًا على الإطلاق. وقد تمَّ تصميمُ منهجيات BPM والتي ستناقَش لاحقاً في هذا الفصل للتأكد من وجود هذه الروابط.

إنَّ عمليـة التخطيط ووَضْع الميزانيـة المالية لديها بنية منطقية تبدأ عـادةً بتلك التكتيكات التي تُولِّد شـكلاً من أشـكال الإيرادات أو الدخل. وتعتمدُ القدرة على توليد الدخل في المنظمات التي تبيع السِّـلع أو الخدمات إمَّا على القدرة على إنتاج السِّـلع والخدمات مباشرة، أو الحصول على الكمية المناسـبة من السـلع والخدمات للبيع. وبعد تحديد رقم الإيرادات؛ فإنه يمكِن توليد التكاليف المرتبطة بتسـليم ذلك المسـتوى من الإيرادات. وفي كثيرٍ من الأحيان؛ فإن ذلك يستلزمُ مدخلات من عِدة أقسام أو تكتيكات. وهذا يعني أنه يجبُ أن تتمَّ هذه العملية بطريقة تعاونية وأن التبعيـة بين المهامِّ تحتاجُ أن يتـمَّ توصيلها وفهمها بوضوح. وإضافةً إلى المدخلات التعاونية؛ فـإن المنظمة تحتاج أيضًـا إلى إضافة العديد من التكاليف العامة؛ فضـلًا عن تكاليف رأس المال المطلـوب. وبمجرد دَمْـج هذه المعلومات؛ فإنهـا تُوضِّح التكلفة بالتكتيك فضـلًا عن المتطلبات النقدية والتمويلية لوضع الخطة موضع التنفيذ.

٣- المراقبة/ التحليل: كيف نؤدي؟

من الضروري مراقبة أداء المنظمة عندما تكون الخطط التشغيلية والمالية قَيْد التنفيذ. وينبغي أن يتناول الإطار الشــامل لرصد الأداء قضيتين رئيســيتين، وهما: ما الذي يجب مراقبته، وكيفية مراقبته. ونظرًا؛ لأنه من المســتحيل ملاحظة كلِّ شيء؛ فإن المنظمة تحتاج إلى التركيز على مراقبة قضايــا مُحدَّدة، وبعد أن تُحدِّد المنظمة المؤشرات أو المقاييس التي يجب ملاحظتها؛ فإنها تحتاج إلى تطوير إستراتيجية لرَصْد تلك العوامل والاستجابة لها بفعالية. وغالبًا ما تُسمَّى هذه المقاييس مــؤشرات الأداء الرئيســية (أو KPIs باختصــار). وســوف نقدِّم نظرةً عامة عــلى عملية تحديد مؤشرات الأداء الرئيســة لاحقًــا في هذا الفصل. وهناك موضوعٌ آخر لــه صِلة باختيار المجموعة المثلى من مؤشرات الأداء الرئيسة وهو طريقة بطاقة نقاط التوازن، والتي سيتمُّ تغطيتها بالتفصيل لاحقًا في هذا الفصل.

٤- الفِعل والضبط: ما الذي نحتاج إلى القيام به بشكلٍ مختلف؟

وســواءً أكانت الشركة مهتمة بتنمية أعمالها أو حتى على الأقل بتحسين عملياتها؛ فإن جميع الإســتراتيجيات تعتمد فعليًّا عــلى مشروعات جديدة كإنتاج منتجات جديدة، أو دخول أســواق جديدة، أو اكتســاب عملاء جُدد أو شركات جديدة، أو تبسيط بعض العمليات. وتتعامل معظم الشركات مع هذه المشــاريع الجديدة بروح من التفاؤل بدلًا من الموضوعية، متجاهلةً حقيقة أن معظم المشاريع القائمة والمشاريع الجديدة تفشل. ولكن ما هي فرصة الفشل؟ من الواضح؛ أن ذلك يعتمد على نوع المشروع (Slywotzky & Weber, 2007).فأفلام هوليود تضع نسبة احتمال للفشل حوالي ٦٠٪. وينطبق الشيء نفسه على عمليات الدَّمج والاستحواذ. كما أن مشاريع تقنية المعلومات الكبيرة تفشل بمعدل ٧٠٪. وبالنسبة للمنتجات الغذائية الجديدة؛ فإن معدل الفشل هو ٨٠٪. أمَّا بالنســبة للمنتجات الصيدلانية الجديدة؛ فإن نســبة احتمال الفشل أعلى من ذلك؛ فهي حوالي ٩٠٪. وعمومًا فإن معدل الفشــل لمعظم المشــاريع القائمة أو المشاريع الجديدة يمتدُّ ما بين ٦٠ و٨٠٪. وبالنظر إلى هذه الأرقام؛ فإن الإجابة على ســؤال «ماذا نحتاج أن نفعل بشــكل مختلف؟» تصبح قضية حيوية.

وتوضِّح الحالة العملية ٣-٥ كيف حققت منظمةٌ كبيرةٌ غير ربحية نتائج استثنائية عن طريق تحويل وتحديث بنية ذكاء الأعمال الخاصة بها مع أجهزة مستودعات البيانات في العصر الحديث.

حالة عملية ٣-٥

AARP تقوم بتحويل بنية ذكاء الأعمال التحتية وتحقق عائد استثمار بنسبة ٣٤٧٪ في ثلاث سنوات

تُعَدُّ AARP, In - وهي الرابطة الأمريكية للأشـخاص المتقاعدين سابقًا - مجموعة ذات شـأن وعضوية مقرُّها الولايات المتحدة، تأسَّسـت في عام ١٩٥٨م من قِبل Ethel Percy Andrus، وهـي مُعلِّمـة متقاعـدة مـن كاليفورنيا حاصلة عـلى الدكتوراه، و Leonard Davis، مؤسـس مجموعة المسـتعمرات Penn لـشركات التأمين. وكما هو مُوضَّـح في موقع الويب الخاص بهم (aarp.org)؛ فـإن الرابطة تُعَدُّ منظمةً غير ربحية وغـير حزبية، كما أنها منظمـةُ رعاية اجتماعية تضمُّ ما يقرب من ٣٨ مليون شـخص؛ إذ تسـاعد الناس على تحويـل أهدافهم وأحلامهـم إلى إمكانيات حقيقيـة، كما تعزِّز المجتمعـات المحلية، وتحارب من أجل القضايا المهمة للعائلات، مثل: الرعاية الصحية، والتوظيف، والأمن، والدخل، والحماية من سوء المعاملة المالية.

تزايد الطلب على ذكاء الأعمال:

في عـام ٢٠٠٢م، أطلقت المنظمة لأوَّل مرة مبادرة ذكاء الأعمال التي من شـأنها أن تمركز المعلومات (تمتلك AARP مكاتب في جميع الولايات الخمسين؛ إضافةً إلى مقاطعة كولومبيا)، وقد تمكَّن موظفوها مع توفر تحليلات حالية وملائمة ودقيقة ومرنة من:

- ملاءمة الخدمات وعروض المنتجات للقاعدة العضوية والتوقعات.

- تحسين ربحية الأعضاء والاحتفاظ بهم، وكَسْبهم.

- حمايـة صـورة العلامة التجارية لـ AARP من خـلال إدارة العلاقات مع مزوِّدي الخدمات التابعين لجهات خارجية.

وقد سـاعدت هذه الرؤية في تعزيز نجاح الرابطة، ومع هذا النجاح؛ جاءت كميات أكبر من البيانات وزيادة الطلب على التحليلات الجديدة.

وبحلـول عـام ٢٠٠٩م، واجه فريـق ذكاء الأعمال تحديًّا كبيرًا؛ إذ لم يَعُد مسـتودع البيانات الخاص بهم - المسـتند على قاعدة بيانات SQL العلائقية من Oracle - قادرًا على مواكبة الطلب؛ إذ شـهد الفريق أكثر من ٣٠ حالة فشـل في تلك السـنة. وقد كان هذا الأمر مكلفًا وغير مقبول.

وإضافــةً إلى ما ســبق؛ فقــد كان أداء النظام مصدر قلقٍ رئيــسي. ونظرًا لتزايد حجم البيانات، وتعَذُّر إكمال التنزيل اليومي إلى مســتودع المســتندات حتى الساعة ٣:٠٠ مساءً - وهــو ما أثر على طول مدة انتظار الموظفين للتقارير. تقول بروني مديرة ممارســة ذكاء الأعــمال، AARP: «كان محللونا يقومــون بإعداد تقرير، ثم الذهــاب لتناول القهوة أو لتنــاول طعام الغداء، وربما يحصلون على الرَّد، إذا كانوا محظوظين، بحلول الســاعة ٥:٠٠ مســاءً، إن هذا الأمر كان غير مقبول؛ فقد كان النظام مشــغولًا للغايــة في كتابة البيانات اليومية الجديدة لدرجة أنه لم يعطِ أيَّ أهمية لعمليات القراءة التي قام بها المستخدمون».

ولم يتمكَّن المحللون أيضًا من إنشــاء استعلامات مُخصَّصة دون تدخل لتقنية المعلومات. فعندمــا تلقَّت تقنية المعلومات طلباً لنوعٍ جديــد من التقارير؛ كان على فريق ذكاء الأعمال تحسين الاستعلامات وإرسال نموذج تقرير مرةً أخرى إلى الطالبين للمراجعة. وتستغرق هذه العمليــة من بدايتها إلى نهايتها من أســابيع إلى شــهور. وقد وجد الموظفــون أخيرًا أنه من المستحيل عمل نسخة احتياطية من النظام كلَّ ليلة بوجود أكثر من ٣٦ تيرابايت من البيانات في مســتودع البيانات. ولذلك؛ فقد اقتصرت النُّسَخ الاحتياطية على بضع جداول حرجة؛ مما جعل من الصعب على الموظفين إنشاء خطة فعَّالة لاستعادة القدرة على العمل بعد الكوارث.

ووفقاً لبروني؛ فإن هذا الأمر إذا تُرِكَ دون حَلٍّ فمن الممكن أن تؤثر هذه التصدُّعات عــلى عمــل الرابطة. تقول بروني: «يوفر التحليل مقاييس أساســية مهمة لتقييم مدى تحقيق عضويتنا وأهدافنا الاجتماعية؛ فمن الضروري تمكين التحســين المســتمر واتخاذ القرارات؛ لدعم احتياجات الأعضاء».

إنشاء بيئة رشيقة لذكاء الأعمال:

نظــرًا لتطلُّع فريق بروني إلى تحديث بيئة ذكاء الأعمال؛ فقد قاموا بتقييم خيارين، وهما: تحســين البيئة الحالية أو الانتقال إلى جهاز مستودع بيانات واحد. تقول بروني: «لقــد وجدنــا تكلفة كلِّ خيــار قابلة للمقارنــة؛ ولكن الجهاز فقط هو الــذي قدَّم لنا نقلــةً نوعيةً فيما يتعلق بالأداء الذي كنَّا نحتاجــه، ومن بين الخيارين المختلفين اللذين درســناهما، استطاع جهاز مســتودعات البيانات IBM Netezza أن يتفوق على الخيار الآخــر؛ نظرًا لكونه أكثر أمانًا؛ إذ إنــه لم يتطلب من نموذج البيانات ضبطًا دقيقًا لمخازن البيانــات الأخــرى. وإضافةً إلى ذلك؛ فقد اســتطعنا تجربة الجهاز قبــل شرائه لمعرفة حقيقة ما إذا كان بإمكانه فِعْل كلِّ ما نحتاج إليه أم لا؛ وذلك على الرغم من أن معظم البائعين لا يقدِّمون هذا النوع من الخيارات وهو (تجربة الجهاز قبل الشراء)».

وقد تبَنَّت المنظمة نموذج تطوير «سكروم» في بناء البيئة الجديدة، والذي يُستخدَم عـادةً مـن قِبل مُطـوِّري البرامج؛ لتوفير إطار عمـل يختـصر دورات التطوير، كما أنه يُسرِّع الوقت لتسـويق طلبات ذكاء الأعمال. وتضيف بروني: «إن استخدام سكروم في مستودعات البيانات؛ هو شيءٌ جديدٌ لم يسمع به أحد؛ لكن المبدأ الأساسي الذي يقدِّمه هو عملية مرنة وتكرارية تمكِّننا من تحويل الاحتياجات التحليلية للمسـتخدِمين بسرعة إلى تقارير تشغيلية تُظهِر بيانات ذات معنى».

وخلال ٩ أشـهر مـن اقتناء الجهاز الجديـد؛ قام الفريق بتحويـل جميع النصوص والإجراءات من قاعدة بيانات أوراكل إلى جهاز مستودع البيانات IBM Netezza. ويتمُّ حاليًّا تنزيلُ مصادر بيانات كلٍّ من الحسابات الأساسية وبيانات العضوية (التي تتواجد في IBM DB2 لقاعدة بيانات z/OS والتي تعمل على خادم IBM System z)، وبيانات الموارد المالية والبشرية من قواعد بيانات أصغر أخرى، وبيانات تحليل وتجزئة الحملات مـن بيانات جهات خارجية في جهاز مسـتودع البيانـات IBM Netezza ليلًا؛ إذ يمكِن الوصول إليها من خلال أدوات ذكاء الأعمال للمنظمة دون انقطاع.

تشغيل استعلامات مُعقَّدة في سرعة البرق:

من حيث الأداء (والذي كان أكثر اهتمامات فريق ذكاء الأعمال إلحاحًا)؛ فإن إكمال تنزيل البيانات اليومية الآن يتمُّ قبل السـاعة ٨:٠٠ صباحًا، وهو تحسن بنسبة ١٤٠٠٪، كما أن إكمال التقارير التي كانت تسـتغرق دقائق من قبل للتشغيل أصبح يتم الآن في عدة ثوان، وهو تحسـن بنسبة ١٧٠٠٪. وفضلًا عن ذلك؛ فقد ساعد اقتناء الجهاز أيضًا في ضغط حجم البيانات من ٣٦ تيرابايت إلى ١,٥ تيرا بايت فقط؛ مما يتيح لفريق العمل الاحتفاظ بنسخة احتياطية من مستودع البيانات في ٣٠ دقيقة فقط.

وعـلى نفس القدر من الأهمية؛ فإن ما يقرب من ٢٢٠ من موظفي الموارد البشرية، والمالية، والتسـويق، إضافةً إلى موظفي الحملة الذين يستخدمون هذا النظام؛ يمكنهم الآن إجـراء مـا تشـير إليه بروني بأنـه «تحليل تدريـب الفكر»؛ بمعنى إنشـاء تقارير مُخصَّصة لاختبـار النظريات المتعَلِّقة باحتياجات العضويـة. وتقول بروني: «إن جهاز مستودع البيانات IBM Netezza يشبه قيادة سيارة فيراري، لقد فتحنا عالمًا جديدًا من الإمكانيـات لعملائنا الداخليين، الذين يمكنهم فعلًا إنشـاءُ تقارير على الفور والحصول عـلى النتائج في غضون ثوانٍ معدودة. وقد شَـهِدنا ارتفاعًا كبـيرًا في عدد التقارير التي

تم إنشـاؤها في الأشهر القليلة الأولى من التشغيل؛ أي ما يقرب من ثلاثة أضعاف الرقم الـذي كنا قد أيَّدنـاه من قبل. وإضافةً إلى الغوص العميق الذي يُمكِنهم القيامُ به الآن؛ فقد شهدنا نموًا ثابتًا في عمليات تجديد العضويات واكتساب أعضاء جدد، والمشاركات».

تحقيق عائد الاستثمار السريع:

كما مكَّـن النظامُ الأسـاسي الجديد المنظمة مـن إعادة نشر موظفـي دَعْم تقنية المعلومـات من مجموعة ذكاء الأعمال إلى مناطق أخرى. فقد كان الفريق في السـابق بحاجةٍ إلى مسـؤول قاعدة بيانات بدوام كامل (DBA) إلى جانب دَعْمٍ جزئي من شبكة منطقة التخزين الخاصة بالمنظمة (SAN)، وفرق الخدمة المتوسطة. تقول بروني: «إنه لأمـرٌ مدهشٌ، لم نَعُد بحاجة إلى دعم تقنية المعلومات؛ إذ يتمُّ شَـحْن جهاز مسـتودع البيانـات IBM Netezza بالفعل. فقط امنحه القوة، واعطه شـبكة، وبفعل ذلك فقد انتهي الأمر، ولا يحتاج أي شيء آخر».

وقد مكَّنت هذه التحسـينات المؤسسـة من تحقيق عائدٍ اسـتثماري بنسبة ٩٪ في السـنة الأولى، مع عائد استثمار متوقع بنسـبة ٢٧٤٪ في السنة الثانية، واستثمار بنسبة ٣٤٧٪ بحلول السنة الثالثة. تقول بروني: «لقد أظهر تحليلنا الأوَّلي عائد استثمار إيجابي في العـام الأول، وهو أمرٌ غير مألوف أبدًا بالنسـبة لتحديث البنية التحتية؛ نظرًا إلى أن جميـع التكاليف غير متكررة في السـنة الأولى، ولقد كان التنفيـذ الفعلي لمعدل العائد على الاستثمار أعلى من ذلك؛ أنهينا عملية المقايضة قبل ثلاثة أشهر من الموعد المحدد».

توسيع تأثير ذكاء الأعمال:

لقـد أدرك فريق بروني القيمـة العالية لذكاء الأعمال في المنظمـة من خلال تحديث بنيتـه التحتيـة. تقول بروني: «بعد أن انتقلنـا إلى IBM Netezza؛ انتشر الكلام بأننا كنا نفعل الأشياء بشـكلٍ صحيح، وأن الاستفادة من الخدمة الداخلية كانت ذكية للغاية؛ لقد اكتسبنا مجالات حيوية جديدة للمهمة، مثل منطقة التأثير الاجتماعي التي تدعم حملاتنا؛ من أجل القضاء على الجوع، وإنشـاء حملات جيدة تستند إلى حقيقة أن لدينا بنية تحتية قوية. وإضافةً إلى ذلك؛ فقد غيَّرنا منهجنا في العمل؛ إذ يمكننا التطوير بطريقة أكثر مرونةً، هـذا من ناحيـة التنمية. أمَّا من ناحيـة إدارة البرنامج؛ فإنها تقلِّـص دورات إطلاقنا من أشهر - والتي تتناسب مع البُنَى التحتية التقليدية لبرمجيات البيانات - فقط لعدة أسابيع.

أسئلة للمناقشة:

١- ما هي التحَدِّيات التي تواجهها الرابطة؟

٢- ماذا كان النهجُ المتَّبع من أجل وَضْع حلٍّ ممكن؟

٣- ما هي النتائج التي تمَّ الحصول عليها على المدى القصير؟ وماذا كانت الخطط المستقبلية؟

Source: IBM customer success story. (2011). AARP transforms its business intelligence infrastructure—Achieving a 347% ROI in three years from BI modernization effort. http://www-03.ibm.com/soft-ware/businesscasestudies/us/en/corp?synkey=A735189Y23828M82 (accessed June 2016).

أسئلة مراجعة على القسم ٣-٩:

١- ما هي إدارة أداء الأعمال؟ كيف ترتبط بذكاء الأعمال؟

٢- ما هي المكونات الرئيسية الثلاثة لنظام BPM؟

٣- اذكر المراحل الأربع لدورة BPM وصِفها بإيجاز.

٤- لماذا تُعَدُّ الإستراتيجية هي الجزء الأكثر أهميةً من تنفيذ BPM؟

٣-١٠ مقياس الأداء:

إن نظام BPM هو نظامٌ لقياس الأداء. ووفقًا لما ذكره Simons (2002)؛ فإنّ نُظُم قياس الأداء؛ هي: «مســاعدة المديرين في متابعة تنفيذ إستراتيجية الأعمال؛ من خلال مقارنة النتائج الفعلية بالأهداف والغايات الإســتراتيجية، وعادةً ما يشتمل نظام قياس الأداء على طرق منهجية لتحديد أهداف العمل مع تقارير ردود الأفعال الدورية التي تشير إلى التقدم في تحقيق الأهداف» (ص ١٠٨).

الجديرُ بالذِّكر أن كلَّ أنظمة القياس تدور حول المقارنات. ونادرًا ما تكون الأرقام الأوَّلية ذات قيمة. فإذا تمَّ إخبارك بأن أحد مندوبي المبيعات قد أكمل ٥٠٪ من الصفقات التي كان يعمل عليها خلال شهر واحد؛ فلن يكون لذلك أيُّ معنى يُذكَر. ولنفترض الآن أنه تمَّ إخبارك بأن الشخص نفسه الذي قام بالمبيعات لديه معدلُ إغلاق شهري بنسبة ٣٠٪ في العام الماضي. فالظاهر أن هذا الاتجاه جيد. ولكن ماذا لو تمَّ إخبارك أيضًا بأن معدل الإغلاق اليومي لجميع مندوبي المبيعات في الشركة كان ٨٠٪؟ وقتها يظهر لك أن مندوب مبيعات معين يحتاج إلى زيادة السرعة. وكما يُوحي تعريف

Simons لقيـاس الأداء؛ فإن المقارنات الرئيسـية تدور حول الإسـتراتيجيات والأهداف والغايات. وعادةً ما تُسمَّى المقاييس التشغيلية المُستَخدَمة لقياس الأداء، مؤشرات الأداء الرئيسية (KPIs).

مؤشر الأداء الرئيسي (KPI):

هناك اختلافٌ بين مقياس «تشغيل مصنع» ومقياس «الانحياز الإستراتيجي»، وغالبًا ما يُستخدَم مصطلـح مؤشر الأداء الرئيسي (KPI) للدلالة على المصطلح الأخير (الانحياز الإسـتراتيجي). يمثل مـؤشرُ الأداء الرئيسي هدفًا إسـتراتيجيًا ويقيس الأداء مقارنةً بالهـدف. ووفقًا لما ذكره Eckerso (2009)؛ فـإن مؤشرات الأداء الرئيسـة تكون متعددة الأبعاد، تحمـل عِدَّة معانٍ، وهذا يعني أن مؤشرات الأداء الرئيسة لها عِدَّة مميزات متنوعة، وتتضمَّن:

- **إستراتيجية:** حيث تجسِّد مؤشرات الأداء الرئيسة هدفًا إستراتيجيًا.
- **أهداف:** فمؤشرات الأداء الرئيسية تقوم بقياس الأداء مقارنةً بالأهداف المُحدَّدة؛ إذ يتمُّ تحديد الأهداف في الإستراتيجية أو التخطيط أو جلسات الميزانية، ويمكن أن تتخذ أشكالًا مختلفة (على سبيل المثال: أهداف الإنجاز، وأهداف التخفيض، والأهداف المطلقة).
- **مجالات:** حيث تحتوي الأهداف على مجالات أداء (على سـبيل المثال: أعلى من الهدف، أو في صميم الهدف، أو أدنى من الهدف).
- **ترميـز:** حيـث يتمُّ ترميز المجـالات في البرامج؛ مما يتيح العرض المرئي للأداء (على سـبيل المثال: الأخضر، والأصفر، والأحمر). ويمكن أن يرتكز الترميزُ على النِّسَب المئوية أو القواعد الأكثر تعقيدًا.
- **الأطر الزمنية:** حيث يتمُّ تعيينُ أُطُر زمنية للأهداف التي يجب إنجازها. وكثيرًا ما يتم تقسـيم الإطار الزمني إلى فواصل زمنية أصغر؛ من أجل الحصول على نقاط ميل لمنحنى الأداء.
- **المعايير:** يتمُّ قياس الأهداف وفقًا لخط الأساس، أو المؤشر المرجعي. وغالبًا ما تكون نتائج العام السابق بمثابة معيار، كما يمكِن أيضًا استخدام أرقام عشوائية أو مؤشرات خارجية.

Sources: Campbell, C. (2015). Top five differences between data lakes and data warehouses. www.blue-granite.com/blog/bid/402596/Top-Five-Differences-between-Data-Lakes-and-Data-Warehouses (accessed July 2016); Woods, D. (2011, July). Big Data requires a big, new architecture. Forbes.www.forbes.com/sites/ciocentral/2011/07/21/big-data-requires-a-big-new-architecture/#598623291d75 (accessed August 2016); Dixon, J. (2010). Pentaho, Hadoop, and data lakes. James Dixon's Blog.https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/(accessed August 2016); Dull, T. (2016). Data lake vs data warehouse: Key differences. KDnuggets.com. http://www.kdnuggets.com/2015/09/data-lake-vs-data-warehouse-key-differences.html (accessed August 2016).

في بعض الأحيان يتمُّ التمييزُ بين مؤشرات الأداء الرئيســة التي يُقصَد بها «النتائج» وتلك التي يُقصَد بها «الدوافع»؛ فنتائج مؤشرات الأداء الرئيسة - التي تُعرَف أحيانًا بمؤشرات التباطؤ - تقيس النتائج التي وصل إليها النشــاط الســابق (مثل: الإيرادات). فهي في الغالب مالية بطبيعتها، غير أن هذا ليس بشكلٍ دائم. وتُحدِّد دوافع مؤشرات الأداء الرئيسة - والتي تعرف أحيانًا بالمؤشرات الرئيســة أو محركات القيمة - الأنشــطة التي لها تأثيرٌ كبيرٌ على مؤشرات الأداء الرئيســة للنتائج (مثل: العملاء المحتملون للمبيعات).

ويُطلَق على مؤشرات الأداء الرئيســية (KPIs) في بعض الدوائر أحيانًا اســم KPIs التشغيلي، وهو من التناقض اللغوي (Hatch, 2008). وتقوم معظمُ المؤسسات بجمع مجموعة واسعة من المقاييس التشغيلية. وكما يُوحِي الاسم؛ فإن هذه المقاييس تتعامل مع الأنشطة التشغيلية وأداء الشركة. وتُوضِّح الأمثلة التالية تنوعَ مجالات التشغيل التي تغطيها هذه المقاييس:

- **أداء العميل:** مقاييس لرضا العملاء، وسرعة ودقة حلِّ المشكلة، والاحتفاظ بالعملاء.
- **أداء الخدمــة:** مقاييــس معدلات دقة اتصــال الخدمة، ومعدلات تجديــد الخدمة، واتفاقيات مستوى الخدمة، وأداء التسليم، ومعدلات الإرجاع.
- **عمليــات المبيعات:** حســابات جديدة على وشــك الحدوث، واجتماعــات المبيعات المضمونة، وتحويل الاستفسارات إلى العملاء المتوقعين، ومتوسط وقت إغلاق المكالمات.
- **خطــة المبيعــات/ التوقعات:** مقاييس دقة أســعار الشراء، ونســبة الشراء إلى معــدل التنفيذ، والكمية المحققة، ونسبة توقعات الخطة، وإجمالي العقود المغلقة.

وسواءٌ أكان المقياسُ التشغيلي إستراتيجيًّا أم لا؛ فإن ذلك يعتمدُ على الشركة ومدى استخدامها للمقياس. وفي العديدِ من الحالات؛ فإن هذه المقاييس تمثل الدوافع الرئيسة للنتائج الإستراتيجية، فعلى ســبيل المثال، تذكر Hatch (2008) حالة موزِّع نبيذ من الدرجة المتوســطة تمَّ تقليصه في المراحــل الأولى من خــلال دَمْج الموزِّدين وتياره مــن خلال دَمْج تجار التجزئــة. وردًّا على ذلك؛ قــرَّرت التركيز على أربعة إجراءات تشــغيلية: تعزيز توفُّر المخزون المتاح/ في الوقت المناســب، وقيمة الطلبات «المفتوحة» المُعلَّقة، والحســابات الجديدة الصافية، وتكاليف الترقية والعائد على الاســتثمار في التســويق. وقد كانت النتيجة الصافية لجهودها زيادة بنسبة ١٢٪ في الإيرادات في عامٍ واحد. ومن الواضح أن هذه المقاييس التشــغيلية كانت مُحرِّكات رئيســة. ومع ذلك؛ وكما هو مُوضَّح في القسم التالي؛ فإن الشركات في كثيرٍ من الحالات، تقوم ببساطة بقياس ما هو ملائم مع الحدِّ الأدنى مع النظر في سبب جمع البيانات. وتكون النتيجة إهداراً للوقت والجهد والمال.

نظام قياس الأداء:

يُوجَد فرق بين نظام قياس الأداء ونظام إدارة الأداء؛ إذ يشتملُ الأخير (نظام إدارة الأداء) على الأول (نظام قياس الأداء). فأيُّ نظام إدارة أداء يحتوي على نظام قياس أداء، والعكس غير صحيح. وإذا كنتَ ستسـأل؛ فإن معظم الشركات اليوم تدعي أنهـا تمتلك نظامًا لقياس الأداء؛ ولكن ليس بالضرورة نظام إدارة أداء، على الرغم من أن نظام قياس الأداء لا يُستخدَم إلا قليلاً جدًا، وإن وُجِد؛ فيكون بدون الهيكل الشامل لنظام إدارة الأداء.

تُعَدُّ أنظمة قياس الأداء الأكثر شـيوعًا المُسـتَخدَمة، هي بعض أشـكال بطاقة الأداء المتوازنة من Kaplan و(BSC) Norton. وتشـيرُ دراسـات المقارنة والاسـتطلاعات المختلفة إلى أن نسبة الشركات التي نفذت شـكلًا من أشكال BSC في وقتٍ أو آخر تتراوح بين ٥٠٪ إلى أكثر من ٩٠٪ في أيِّ مـكان. وعلى الرغم من أن هناك بعض الغموض حول ما يُشـكل (التوازن)، غير أنه لا يُوجَد أدنى شـك حول مُنشِـئَيْ BSC؛ وهما: Kaplan و(1996) Norton، واللذين ذكرا أن: «المركزية لمنهجية BSC هي رؤية شـاملة لنظام القياس المرتبط باتجاه المنظمة الإستراتيجي. والتي تستند إلى أربع وجهات نظر من منظور العالم، مع تدابير مالية ذاتية، ومدعومة من قِبل العملاء؛ إضافةً إلى مقاييس التعلُّم والنمو».

أسئلة مراجعة على القسم ٣-١٠:

١- ما هي نُظُم إدارة الأداء؟ ولماذا نحتاج إلى أحدها؟

٢- ما هي السِّمات المميِّزة لمؤشرات KPI؟

٣- اذكر مع التوضيح بإيجاز مجالات التشغيل الأربعة الأكثر شيوعًا لمؤشرات الأداء الرئيسية.

٤- ما هو نظام قياس الأداء؟ وكيف يعمل؟

٣-١١ بطاقات الأداء المتوازنة:

لعلَّ أوسـع أنظمة إدارة الأداء شـهرةً وأكثرها استخدامًا على نطاقٍ واسـع؛ هو بطاقة الأداء المتـوازن (BSC). وقد أوضـح كلٌّ من Kaplan وNorton هذه المنهجيـة لأول مرة في مقالهما -المنشور في Harvard Business Review عام ١٩٩٢م - وكان عنوان هذه المقالة: «بطاقة الأداء المتوازن: المقاييس التي تدفع الأداء». وبعد مرور بضع سـنوات، وتحديدًا في عام ١٩٩٦م، أصدر هذان المؤلفان كتابًا رائدًا بعنوان - بطاقة الأداء المتوازن: ترجمة الإسـتراتيجية إلى فِعل - والذي

وثَّــق كيف كانت الشركات تســتخدمُ BSC ليس فقط لتكمِّل إجراءاتهــا المالية مع المقاييس غير المالية، ولكن أيضًا لتوصيل وتنفيذ إســتراتيجياتها. وعلى مدى السنوات القليلة الماضية، أصبحت BSC مصطلحًــا عامًّا يُســتَخدَم لتمثيل كل نوع مــن أنواع تطبيــق الأداء والتنفيذ، بغض النظر عــمَّا إذا كان متوازنًا أو إســتراتيجيًّا. وردًّا على هذا الاختلاف حــول المصطلح؛ فقد أصدر كلٌّ من Kaplan وNorton كتابًــا جديدًا في عام ٢٠٠٠م، بعنوان: المنظمة التي تركِّز على الإســتراتيجية: كيف تزدهر شركات بطاقات الأداء المتوازن في بيئة الأعمال الجديدة، وقد تمَّ تصميم هذا الكتاب لإعادة التأكيد على الطبيعة الإستراتيجية لمنهجية BSC. وتبع ذلك ببضع سنوات، في عام ٢٠٠٤م، بخرائــط الإســتراتيجية: تحويل الأصول غير الملموســة إلى نتائج ملموســة، والتــي تَصِف عملية مُفصَّلة لربط الأهداف الإســتراتيجية بالتكتيكات والمبادرات التنفيذية. وأخيرًا؛ جاء كتابهم الأخير The Execution Premium، والذي نُشِر في عام ٢٠٠٨م، ويركِّز على الفجوة الإســتراتيجية التي تربط بين صياغة الإستراتيجية والتخطيط والتنفيذ التشغيلي.

وجهات النظر الأربعة:

تشيرُ بطاقة النتائج المتوازنة إلى أننا ننظر إلى المنظمة من أربعة منظورات، وهي:

١- العملاء.

٢- العمليات المالية.

٣- عمليات الأعمال الداخلية.

٤- التعلم والنمو.

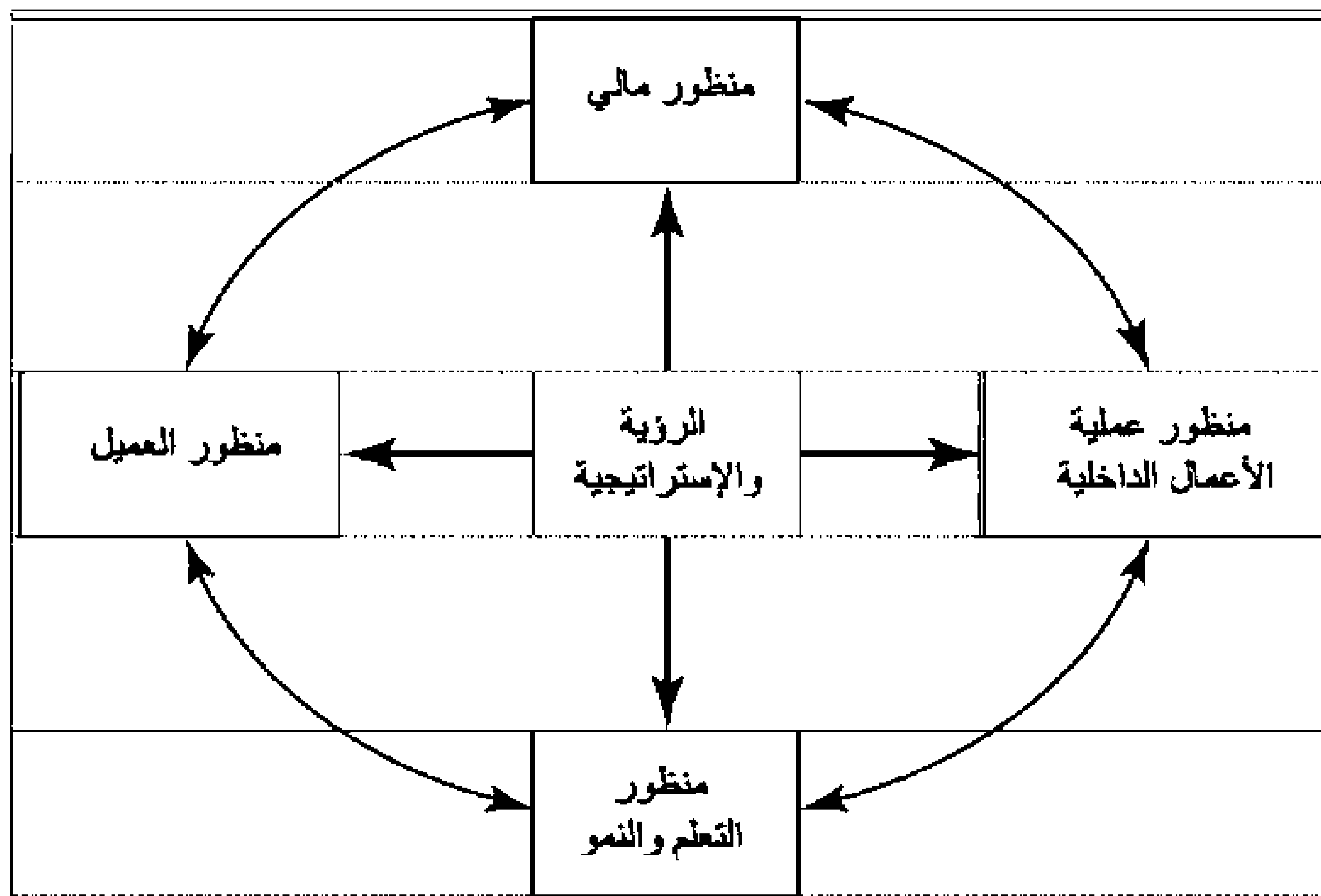
إضافةً إلى تطوير الأغراض، والمقاييس، والأهداف، والمبادرات المتعلقة بكل منظور من هذه المنظورات، ويُبيِّن الشكل ٣-١٣ هذه المنظورات الأربعة وعلاقاتها المتداخلة مع رؤية المنظمة وإستراتيجيتها.

- **منظور العملاء:** أدركت فلسفاتُ الإدارة الحديثة الأهمية الكبيرة للتركيز على العميل ومدى رضاه في أيِّ عملٍ تجاري. وفيما يلي المؤشرات الرئيسة لقياس ذلك؛ فالعملاء إذا لم يشعروا بالرضا؛ فإنهم في الواقع سوف يبحثون عن مورِّدين آخرين يمكنهم تلبية احتياجاتهم. وبالتالي؛ فإن الأداء الضعيــف من هذا المنظور يُعَدُّ مؤشرًا قويًّا على التراجع في المســتقبل، على الرغم من أنَّ الصورة الماليــة الحالية قد تبدو جيــدةً. وفي تطوير المقاييس لتحقيق الرضــا؛ ينبغي تحليل العملاء من حيث أنواع العملاء وأنواع العمليات التي نقدِّم لها منتجًا أو خدمةً لتلك المجموعات من العملاء.

- **المنظور المالي:** لم يتجاهل كلٌّ من Kaplan وNorton الاحتياجات التقليدية للبيانات المالية. ولذلك؛ فإنَّ الأولوية دائمًا ستكون لتقديم بيانات التمويل بشكلٍ دقيقٍ، وفي الوقت المناسب، وسوف يبذل المديرون كلَّ ما في وسعهم لتوفير هذه البيانات. وفي الواقع؛ فإنه غالبًا ما يكون هناك أكثرُ من معالجة كافية لتجهيز البيانات المالية. ومع تنفيذ قاعدة بيانات الشركات؛ فإنه من المنتظر أن تتمَّ معالجة المزيد من البيانات المالية وبطريقة تلقائية. لكن تبقى هناك نقطة مهمة في هذا الأمر، وهي أن التركيزَ الحالي على الأمور المالية يؤدي إلى الوَضْع «غير المتوازن» فيما يتعلق بالمناظير الأخرى. فقد تكون هناك حاجةٌ لإدراج بيانات إضافية مرتبطة بالقطاع المالي في هذه الفئة، مثل: بيانات تقييم المخاطر والفوائد من حيث التكلفة.

- **منظور التعلم والنمو:** يهدف هذا المنظور إلى الإجابة عن سؤال «كيف سنحافظ على قدرتنا على التغيير والتحسين؛ من أجل تحقيق رؤيتنا؟». وهذا يتضمَّن تدريب الموظفين وإدارة المعرفة والخصائص الثقافية للشركات المتعلقة بالتحسينات الفردية والجماعية على حَدٍّ سواء. وفي المناخ الحالي للتغيُّر التقني السريع؛ أصبح من الضروري أن يكون العاملون في مجال المعرفة في وَضْع التعلُّم والتطور المستمر. ويمكن وَضْع مقاييس لتوجيه المديرين إلى تركيز الموارد المالية المخصصة للتدريب؛ إذ يمكِنهم المساعدة أكثر. وعلى أية حال؛ فإن التعلم والنمو يشكِّل الأساس الجوهري لنجاح أيِّ منظمة عاملة في مجال المعرفة. ويؤكد كلٌّ من Kaplan وNorton أن «التعلم» أكثر من كونه «تدريباً»؛ فهو إضافةً إلى التدريب يتضمَّن أيضًا أشياء مثل وجود الموجِّهين والمعلمين داخل المنظمة؛ إضافةً إلى سهولة التواصل بين العمال؛ مما يسمح لهم بالحصول على مساعدة بشأن أيِّ مشكلة عند الحاجة إلى ذلك.

- منظور عملية الأعمال الداخلية: ويركِّز هذا المنظور على أهمية العمليات التجارية؛ إذ تسمح المقاييس التي تستند إلى هذا المنظور للمديرين بمعرفة مدى جودة تشغيل عملياتهم ووظائفهم الداخلية، وما إذا كانت نتائجُ هذه العمليات (أيْ: المنتجات والخدمات) تلبِّي وتتجاوز متطلبات العملاء (المهمة).



شكل ٣-١٣: أربع وجهات نظر في منهجية بطاقة النتائج المتوازنة

معنى التوازن في BSC:

من وجهة نظر عالية المستوى؛ فإن بطاقة الأداء المتوازن (BSC) تُعَدُّ مقياسًا للأداء ومنهجيةَ إدارة تساعد في ترجمة الأوضاع المالية للمؤسسة، والعملاء، والعمليات الداخلية، إضافةً إلى أهداف التعلُّم والنموِّ إلى مجموعةٍ من المبادرات القابلة للتطبيق. وباعتبار بطاقة الأداء المتوازن (BSC) منهجيةً للقياس؛ فقد تمَّ تصميمها لتجاوز القيود المفروضة على الأنظمة التي تركِّز على الناحية المالية. وتقوم هذه المنهجية بذلك عن طريق ترجمة رؤية المنظمة وإستراتيجيتها إلى مجموعة من الأهداف والتدابير والمبادرات المالية وغير المالية المترابطة.

وتندرجُ الأهداف غير المالية في واحدةٍ من ثلاث وجهات نظر:

- **العميل:** يُحدِّد هذا الهدف كيف يجبُ أن تظهر المنظمة لعملائها إذا أرادت تحقيق رؤيتها.
- **عملية الأعمال الداخلية:** ويُحدِّد هذا الهدف العمليات التي يجبُ أن تتفوق فيها المؤسسة لإرضاء مساهميها وعملائها.
- **التعلُّم والنمو:** يوضِّح هذا الهدف كيف يمكن لمنظمةٍ ما تحسين قدرتها على التغيير والتحسين؛ من أجل تحقيق رؤيتها.

وفي الأســاس؛ فإن الأهداف غير المالية تشكِّل سلسلةً سببيةً بسيطةً مع «التعلم والنمو»؛ مما يــؤدي إلى تغيير «عمليــة العمليات التجارية الداخلية»، والتي تنتج نتائج «العميل» والمســؤولة عــن تحقيق الأهداف «المالية» للشركة. وفي BSC، ينشــأ مصطلح التوازن؛ لأن مجموعة التدابير مجتمعة من المفترض أن تشمل المؤشرات:

- المالية وغير المالية.
- الرائدة والمتخلفة.
- الداخلية والخارجية.
- الكمية والنوعية.
- على المدى القصير والمدى الطويل.

أسئلة مراجعة على القسم ٣-١١:

١- ما هي بطاقة الأداء المتوازن (BSC)؟ ومن أين تأتي؟

٢- ما هي المناظير الأربعة التي تقترحها (BSC) لعرض الأداء التنظيمي؟

٣- لمــاذا نحتاج إلى تحديدِ مقاصد ومقاييس وأهداف ومبادرات منفصلة لكلٍّ من وجهات النظر الأربعة هذه؟

٤- ما هو معنى التوازن في BSC؟ وما هو تحفيز ذلك التوازن؟

٥- ما هي أوجه الشبه والاختلاف بين لوحات المعلومات وبطاقات الأداء؟

٣-١٢ ستة سيجما كنظام لقياس الأداء:

منذ إنشاء ستة سيجما في منتصف الثمانينيات، وهي تتمتَّع بانتشارٍ واسع النطاق بين الشركات في جميع أنحاء العالم. ومن الجدير بالذكر؛ أن معظم الشركات لم تستخدمها كمنهجية قياس الأداء والإدارة؛ بل أُستُخدِمَتْ بدلًا من ذلك كمنهجيةٍ لتحسين العمليات؛ إذ تُمكِّن الشركات من التدقيق في عملياتها وتحديد المشــكلات وتطبيق العلاجات. وفي الســنوات الأخيرة أقرَّت بعض الشركات، مثل Motorola بقيمة اســتخدام ستة سيجما لأغراض إســتراتيجية؛ إذ توفر ستة سيجما في هذه الحالة الوســائل لقياس ومراقبة العمليات الأساسية المتعلقة بربحية الشركة، وتسريع التحسُّن في الأداء العام للأعمال. ونظرًا لتركيزها على العمليات التجارية؛ فإن ســتة سيجما توفِّر أيضًا طريقةً مباشرةً لمعالجة مشكلات الأداء بعد تحديدها أو اكتشافها.

تُعرَّف سـيجما (σ) Sigma بأنها حرفٌ في الأبجدية اليونانية يسـتخدمه الإحصائيون لقياس التبايـن في العمليـة. وفي مجال الجودة يكـون التغيُّر مرادفًا لعدد العيوب. وبشـكلٍ عام؛ فقد قبِلـت الشركات قدرًا كبيرًا من التبايـن في عملياتها التجارية. وأمَّا من الناحية العددية؛ فقد كان المعيار ٦٢٠٠ إلى ٦٧٠٠٠ عَيْب لكلِّ مليون فرصة (DPMO). فعلى سبيل المثال: إذا كانت شركة تأمين تتعامل مع مليون شـكوى؛ فإن إجراءات التشـغيل العادية من ٦٢٠٠ إلى ٦٧٠٠٠ من هذه الشكاوى تُعَدُّ معيبةً (أيْ: يتمُّ التعامل معها بشكل غير صحيح، وتحتوي على أخطاء في النماذج). يمثل هذا المسـتوى من التباين مسـتوى أداء من ثلاثة إلى أربعة Sigma. ولتحقيق مستوى أداء سـتة Sigma؛ فإنه سـيكون على الشركة تقليـل عدد العيوب إلى ما لا يزيـد عن ٣٫٤ DPMO. ولذلك؛ فإن ستة سيجما عبارة عن منهجية لإدارة الأداء تهدف إلى تقليل عدد العيوب في العمليات التجارية إلى أقرب ما يكون إلى الصفر من DPMO قدر الإمكان.

نموذج الأداء DMAIC:

تقع سـتة سيجما على نموذجٍ بسـيط لتحسين الأداء يُعرَف باسـم DMAIC. ويشبه نموذج DMAIC هـذا نمـوذج BPM؛ فكلاهما نموذج لتحسـين الأعمال في حلقة مُفرَّغة، وهو يشـمل خطـوات تحديـد وقياس وتحليل وتحسـين والتحكُّم في العملية. ويمكـن وَصْف الخطوات على النحو التالي:

١- **تعريف:** ويُقصَد بذلك تحديد مقاصد وأهداف وحدود نشاط التحسين. وعلى المستوى الأعلى؛ فإن الأهداف يُقصَد بها الأهداف الإسـتراتيجية للشركة. وفي المسـتويات الأدنى (مسـتويات الإدارة أو المشروع)؛ فإن الأهداف تركز على عمليات تشغيلية مُحدَّدة.

٢- **قيـاس:** ويُقصَـد به قيـاسُ النظام الحالي. بمعنى وَضْع مقاييس كمية من شـأنها أن تُسـفِر عن بيانات صحيحة إحصائيًّا؛ بحيث يمكِن اسـتخدام البيانـات؛ لمراقبة التقدُّم نحو الأهداف المُحدَّدة في الخطوة السابقة.

٣- **تحليل:** ويُقصَد به تحليلُ النظام لتحديد سُبُل القضاء على الفجوة بين الأداء الحالي للنظام أو العملية والهدف المنشود.

٤- **تحسـين:** ويُقصَـد به الشروع في إجـراءات للقضاء على الفجوة من خـلال إيجاد طرقٍ لفعل الأشـياء بشـكلٍ أفضل وأسرع وأقل تكلفة. ويتمُّ ذلك باسـتخدام إدارة المشـاريع وأدوات التخطيط الأخرى لتنفيذ النَّهج الجديد.

٥- التحكُّــم: ويُقصَد بذلك إضفاءُ الطابع المؤســسي على النظام بعــد تعديله؛ من خلال تعديل أنظمة التعويض والحوافز والسياسات والإجراءات، وتخطيط موارد التصنيع وموارد الميزانيات وتعليمات التشغيل أو أنظمة الإدارة الأخرى.

وبالنســبة للعمليات الجديدة؛ فإن نموذج DMADV الــمُستخدَم لـــ (التعريف، والقياس، والتحليــل، والتصميــم، والتحقيق) يُعَدُّ نموذجًا تقليديًّا. ولذلك؛ فقد تمَّ اســتبداله بـ DMAIC، غــير أن DMADV لا يزال يُســتخدَم في المقام الأول مع القضايا التشــغيلية، ومع ذلك؛ فلا شيء يمنــع تطبيق هذه المنهجيات. ففي الســنوات الأخيرة، كان هنــاك تركيزٌ على الجمع بين منهجية ســتة ســيجما وغيرها من المنهجيات الناجحة. فمثــلًا؛ تمَّ دَمْج المنهجية المعروفة باســم Lean Manufacturing أو Lean Production أو ببســاطة باســم Lean مع ســتة ســيجما؛ لتحسين تأثيرها في إدارة الأداء.

بطاقة الأداء المتوازن مقابل ستة سيجما:

عــلى الرغم من قيام الكثيرين بالجمع بين ســتة ســيجما وبطاقات قيــاس متوازنة؛ من أجل الحصول على حلٍّ أكثر شــمولية؛ فــإنَّ بعضهم ركَّز على تفضيل إحداها عــلى الأخرى. وقد قدَّم Gupta (2006) في كتــابٍ لــه بعنــوان: « ســتة ســيجما Business Scorecard» ملخصًا جيدًا للاختلافات بين بطاقة الأداء المتوازنة، ومنهجيات ســتة ســيجما (انظر: الجدول ٣-٧). وباختصار؛ فإن الفرق الرئيس يتمثل في أنَّ BSC تركِّز على تحســين الإســتراتيجية العامة؛ في حين تركِّز ســتة سيجما على تحسين العمليات.

جدول ٣-٧: مقارنة بين بطاقة الأداء المتوازن وستة سيجما

بطاقة قياس الأداء المتوازن	ستة سيجما
نظام الإدارة الإستراتيجية	نظام قياس الأداء
تتعلَّق بالعرض الطويل الأجل للأعمال	تقدِّم لقطــةً لأداء الأعمال، وتُحــدِّد التدابير التي تدفع الأداء نحو الرِّبحية
مُصمَّمة لتطوير مجموعة متوازنة من التدابير	مُصمَّمة لتحديد مجموعة من القياسات التي تؤثر على الرِّبحية

بطاقة قياس الأداء المتوازن	ستة سيجما
تُحدِّد القياسات حول الرؤية والقيم	تُحـدِّد المساءلة عن القيادة؛ من أجل الصحة والرِّبحية
تتمثل عمليات الإدارة الحاسمة في توضيح الرؤية/ الإستراتيجية، والتواصل، والتخطيط، وتحديد الأهداف، ومواءمة المبادرات الإستراتيجية، وتعزيز التغذية المرتدَّة	تشمل جميع العمليات التجارية: الإدارية والتشغيلية
أرصدة العملاء والعمليات الداخلية دون دورٍ قياديٍّ واضح المعالم	أرصدة الإدارة وأدوار الموظفين، وأرصدة تكاليف وعائدات العمليات الثقيلة
تؤكِّد على تحقيق كمية مبيعات مُستَهدَفة لكل قياس	تؤكِّد على مُعدَّل التحفيز القوي لكلِّ قياس؛ بغضِّ النظر عن تحقيق كمية المبيعات المُستَهدَفة
تؤكِّد على تعلُّم المديرين التنفيذيين على أساس التغذية المرتدَّة	تؤكِّد على التعلم والابتكار على جميع المستويات بناءً على التغذية المرتدَّة، ويستلزم مشاركة جميع الموظفين
تركِّز على النمو	تركِّز على تعظيم الرِّبحية
ضخمة بالنسبة للمحتوى الإستراتيجي	ضخمة عند التنفيذ للربحية
نظام إدارة يتكوَّن من التدابير	نظام القياس على أساس إدارة العملية

Source: Gupta, P. (2006). Six Sigma business scorecard, 2nd ed. New York: McGraw-Hill Professional.

قياس الأداء الفعَّال:

يقـدِّم عددٌ من الكتب الطرق التي تُحـدِّد ما إذا كانت مجموعةٌ من مقاييس الأداء جيدةً أم سيئة. ومن بين الصفات الأساسية التي تميِّز مجموعة مقاييس الأداء الجيدة ما يلي:

- ينبغي أن تركِّز المقاييس على العوامل الرئيسية.
- يجب أن تكون المقاييسُ مزيجًا من الماضي والحاضر والمستقبل.
- يجب أن تُوازن المقاييس بين احتياجات المساهمين والموظفين والشركاء والمورِّدين وأصحاب المصلحة الآخرين.

- يجب أن تبدأ الإجراءات من الأعلى، ثم تتدفق إلى الأسفل.
- يحتاجُ القياس أن يكون له كمية مبيعات مُستَهدَفة (هدف يسعى إلى تحقيقه) والتي تستند إلى البحث والواقع وليست تعسُّفية.

كما يُلاحِظ القسم المتعلِّق بمؤشرات الأداء الرئيسية أنه على الرغم من أهمية جميع هذه الخصائص؛ فإن المفتاح الحقيقي لنظام قياس الأداء الفعَّال هو وجود إستراتيجية جيدة. فيجبُ اشتقاق المقاييس من إستراتيجيات وحدة الشركات والأعمال، ومن تحليل العمليات التجارية الرئيسية المطلوبة لتحقيق تلك الإستراتيجيات. وبطبيعة الحال؛ فالقول أسهل من الفعل. فلو كانت بسيطةً؛ فإن معظم المنظمات لديها بالفعل أنظمةٌ فعَّالةٌ لقياس الأداء، غير أنها لا تقوم بذلك.

وتقدِّم الحالة العملية (٣-٦) - والتي تصِف نظام بطاقة الأداء على أساس KPI المستند إلى الويب على Expedia.com - نظرةً ثاقبةً على الصعوبات المتعلقة بتحديد كلٍّ من نتائج ومؤشرات KPI، وأهمية مواءمة مؤشرات الأداء الرئيسة للإدارات مع الأهداف العامة للشركة.

حالة عملية ٣-٦

بطاقة درجة رضا العملاء الخاصة بـ Expedia.com

تُعَدُّ شركة Expedia, Inc. هي الشركة الأم لبعض الشركات العالمية الرائدة في مجال السياحة؛ إذ تقدِّم خدمات السفر للمسافرين بغرض الترفيه، والمقر الرئيسي لهذه الشركة هو الولايات المتحدة الأمريكية، كما أنَّ لها فروعًا أخرى في مختلف دول العالم. وهي تمتلك وتدير مجموعةً متنوعةً من العلامات التجارية المعروفة، بما في ذلك، Expedia.com، وHotels.com، وHotwire.com، وTripAdvisor، وEgencia، وClassic Vacations، ومجموعة من الشركات المحلية والدولية الأخرى. وتحتوي عروض سفر الشركة على: رحلات شركات الطيران، والإقامة في الفنادق، واستئجار السيارات، وخدمات أماكن الوصول، والرحلات البحرية، ورحلات السفر التي تقدِّمها مختلف شركات الطيران؛ إضافةً إلى أماكن الإقامة، وشركات تأجير السيارات، والخدمات التي تُقدَّم في مكان الوصول، وخطوط الرحلات البحرية، وغيرها من شركات خدمات السياحة القائمة بذاتها أو التي تعمل في مجموعة. كما أنها تُسهِّل حجز غرف الفنادق ومقاعد شركات الطيران وتأجير السيارات، وخدمات الوصول المقدَّمة من الشركات السياحية. وتعمل هذه الشركة كوكيلٍ في الإجراءات؛ إذ تقوم بالحجز للمسافرين على شركة الطيران أو الفنادق أو شركات تأجير السيارات أو خطوط الرحلات. وتجعل هذه العلامات التجارية المشهورة

مع الأعمال التجارية المبتكرة من Expedia أكبر وكالة سـفر عبر الإنترنت في العالم، وهي ثالث أكبر شركة سفر في الولايات المتحدة، ورابع أكبر شركة سفر في العالم. وتتمثل مهمتُها في أن تصبح أكبرَ وأكثر شركة سـياحية مربحةً في العالم؛ وذلك من خلال مساعدة الجميع في كلِّ مكان في التخطيط وتوفير كلِّ ما يتعلق بخدمات السياحة والسفر.

المشكلة:

يُعَـدُّ رضا العملاء هـو المِفتاح لكلِّ مهامٍّ وإسـتراتيجيات؛ بل وأسـاس نجاح شركة Expedia. ونظرًا لأن خدمة البيع بالتجزئة هي نشاطٌ تجاريٌّ عبر الإنترنت؛ فإن تجربة تسـوق العميل تُعَدُّ أمرًا بالغ الأهمية لإيرادات شركة Expedia؛ إذ إن تجربة التسـوُّق عبر الإنترنت يمكنها أن تجعل النشاط التجاري عبر الإنترنت يزدهر أو ينتهي، ومن المهمِّ أيضًا أن تُتوَّج تجربة التسـوق للعميل برحلة جيدة. ونظرًا لأن تجربة العميل تُعَدُّ أمرًا بالغ الأهمية؛ فإنه يجبُ تتبُّع جميع مشـكلات العملاء ومراقبتها وحلُّها في أسرع وقتٍ ممكن. وللأسـف؛ فقد كانت شركة Expedia قبل بضع سنوات لا تستطيع الوصول إلى «رأي العميـل». فلم يكن لديهـا طريقة موحَّدة لقياس مدى رضـا العميل، أو تحليل محركات الرضا، أو تحديد أثر الرضا على ربحية الشركة أو أهداف العمل العامة.

الحل:

لم تكن مشكلة شركة Expedia تتمثل في نقص البيانات. أدركت المجموعة المسئولة عن قياس مدى رضا العملاء في شركة Expedia أن لديها الكثير من البيانات. وبشـكلٍ عـام؛ فقد كان هنـاك ٢٠ قاعدة بيانات متباينة مع ٢٠ مالـكًا مختلفًا. ومن ثمَّ قامت المجموعة بتكليف أحد مُحلِّلي أعمالها بمهمة تجميع البيانات من هذه المصادر المختلفة وتجميعها في عددٍ من المقاييس الأساسية لقياس مدى الارتياح. وقد كان محلل الأعمال يقضي ما يقرب من أسـبوعين إلى ثلاثة أسـابيع كلَّ شهر في سـحب البيانات وتجميعها، وهـذا الأمر لم يترك وقتًـا للتحليل نهائيًا. ولذلك؛ فقد أدركـت المجموعة في النهاية أن مجـرد تجميع البيانات لم يكن كافيًا؛ إذ يجب أن يُنظر إلى البيانات في سـياق الأهداف الإستراتيجية، وكان يجب على الأفراد أن يضعوا أيديهم على النتائج.

ولمعالجة المشـكلة قرَّرت المجموعة أنها بحاجةٍ إلى رؤيةٍ راقية. ولذلك فقد بدأت بتحليل مُفصَّل للدوافع الأساسية لأداء الإدارة والرابط بين هذا الأداء وأهداف Expedia الشـاملة. ثم قامت المجموعة بعد ذلك بتحويل برامج التشغيل وهذه الارتباطات إلى بطاقة قياس أداء. وقد تضمَّنت هذه العملية ثلاث خطوات:

١- **تقرير كيفية قياس مدى رضا العميل:** وقد تطلَّب ذلك من المجموعة تحديد أيٍّ من المقاييس في قواعد البيانات العشرين سـيكون مفيدًا في إظهار مستوى رضا العميل، وقد أصبح هذا هو الأساس لبطاقات الأداء ومؤشرات الأداء الرئيسية.

٢- **تحديـد أهـداف الأداء الصحيحة:** وهذا يتطلب من المجموعـة تحديدَ ما إذا كانـت أهـداف KPI لها أرباحٌ قصيرة الأجل أو طويلة الأجـل؛ فمجرد رضا العميل عن تجربته عبر الإنترنت لا يعني بالضرورة أن العميل راضٍ عن مُقدِّم خدمة السفر.

٣- **وَضْـع البيانات في السـياق:** فقد كان على المجموعة رَبْـط البيانات بمدى رضا العملاء في الوقت الحالي.

ويتمُّ تغذية مصادر البيانات المختلفة بشكلٍ فوري في قاعدة بيانات رئيسية (تُسمَّى نظام دعم القرار). وفي حالة مجموعة الاستبيانات الخاصة بمدى رضا العملاء؛ فإن هذه الاستبانات تتضمَّن استطلاعات العملاء، وأنظمة إدارة علاقات العملاء، وأنظمة الاستجابة الصوتيـة التفاعلية، وأنظمة خدمة العملاء الأخرى. ويتمُّ تنزيل البيانات في مصنع DSS بشـكلٍ يوميٍّ في صورة عِدَّة مكعبات DMs متعددة الأبعاد؛ بحيث يمكِن للمستخدِمين الوصول إلى البيانات بطرق متنوعة ذات صِلة باحتياجاتهم التجارية الخاصة.

الفوائد:

وفي نهاية المطاف اسـتطاعت المجموعةُ الخاصة من خلال قياس مدى رضا العملاء؛ أن تحقق من ١٠ إلى ١٢ هدفًا ترتبط مباشرةً بمبادرات شركة إكسبيديا. وقد تمَّ رَبْط هذه الأهـداف بدورها بأكثر من KPI 200 ضِمن المجموعة الخاصة بقياس مدى رضا العملاء. يستطيع مالكو KPI بناء وإدارة واستهلاك بطاقات الأداء الخاصة بهم، كما يملك المديرون التنفيذيون ومديرو الإدارة رؤيةً شفافةً حول مدى توافق الإجراءات مع الإستراتيجية. كما توفـر بطاقة النتائج للمجموعة الخاصة بقياس مدى رضا العملاء، القدرةَ على التعمُّق في البيانات التي تكمُنُ وراء أيٍّ من الاتجاهات أو الأنماط الملاحظة. وكلُّ هذا كان يستغرق في الماضي أسابيع أو شهور للقيام به؛ وذلك إذا تم القيامُ به على الإطلاق. ويمكن لمجموعة خدمة العملاء معرفة مدى نجاحها في ما يتعلق بمؤشرات الأداء الرئيسية من خلال بطاقة النقاط، والتي تنعكسُ بدورها على أهداف المجموعة وأهداف الشركة.

كـما أنَّ للبيانات الموجـودة في النظام ميزةً إضافيةً؛ فهـي لا تدعم فقط مجموعة رضا العملاء؛ بل إنها أيضًا تدعمُ وحدات الأعمال الأخرى في الشركة. فعلى سبيل المثال:

يمكِن لمدير الخط الأمامي تحليلُ نفقات شركات الطيران على أساس كلِّ سوق على حِدَة، لتقييم أداء العقد المتفــاوض عليه أو تحديد إمكانات التوفير لتوحيد الإنفاق مع شركة نقل واحدة. كما يُمكِن لمدير الســفر الاستفادة من ذكاء الأعمال لاكتشاف المناطق التي بها كميات كبيرة من التذاكر غير المُستَخدَمة أو الحجوزات غير المتصلة بالإنترنت ووَضْع إستراتيجيات لضبط السُّلوك وزيادة التوفير العام.

أسئلة للمناقشة:

١- مَــنْ هم عمــلاء Expedia.com؟ ولماذا يُعَدُّ رضا العمــلاء جزءًا مهمًّا جدًا من نشاطهم التجاري؟

٢- كيف ساهم نظام البيع بالتجزئة في تحسين رضا العملاء باستخدام بطاقات الأداء؟

٣- ما هي التحَدِّيات؟ وما هو الحلُّ المقترح؟ وما هي النتائج التي حصلوا عليها؟

Sources: Based on Microsoft. (2005). Expedia: Scorecard solution helps online travel company measure the road to greatness. download.microsoft.com/documents/customerevidence/22483_Expedia_Case_Study.doc (accessed June 2016); Editor's note.(2004). Expedia incorporates customer satisfaction feedback and employee input to enhance service and support. Quirk's Marketing Research Media. http://www.quirks.com/articles/a200420041001/.aspx (accessed July 2016).

أسئلة مراجعة على القسم ٣-١٢:

١- ما هي ستة سيجما؟ وكيف يتمُّ استخدامُها كنظام لقياس الأداء؟

٢- ما هو DMAIC؟ اذكر وَصْفاً بإيجاز لخطوات DMAIC.

٣- قارن بين BSC، وستة سيجما كنظامين متنافسين لقياس الأداء.

٤- ما هي المكوِّنات اللازمة لنظام فعَّال لإدارة الأداء؟

تلخيصٌ لأهمِّ نقاط الفصل:

- يُعرَّف مستودع البيانات بأنه عبارة عن مستودع بيانات تمَّ إنشاؤه بشكلٍ خاص؛ إذ يتمُّ تنظيم البيانات بحيث يمكن للمستخدِمين النهائيين الوصولُ إليها بسهولة لعِدَّة تطبيقات.
- تحتوي DMs على بيانات حول موضوعٍ واحد (كالتســويق مثلًا). يمكن أن يكون DM نســخةً

مـن مجموعـةٍ فرعية من البيانات في مسـتودع البيانات. وتُعَدُّ DMs حـلًّا أقل تكلفةً يمكِن استبداله أو يمكِن أن يكمل مستودع البيانات. كما يمكِن أن تكون DMs مستقلةً عن أو تعتمد على مستودع البيانات.

- يُعَـدُّ نظـام ODS نوعًا من قاعدة بيانات ملفات معلومات العملاء التي يتمُّ اسـتخدامها غالبًا كمنطقة تقسيم لمستودع البيانات.
- يشـملُ تكامـلُ البيانات ثلاث عمليات رئيسـية، هي: الوصول إلى البيانـات، واتحاد البيانات، والتقـاط التغيـير. فعندما يتمُّ تنفيذ هذه العمليات الثلاث بشـكلٍ صحيح؛ يمكِن الوصولُ إلى البيانات وإتاحتها لمجموعة من ETL، وأدوات التحليل وبيئات مستودعات البيانات.
- تقوم تقنيات ETL بسَحْب البيانات من مصادر عديدة، وتطهيرها، وتنزيلها في مستودع بيانات. وتُعَدُّ ETL عمليةً متكاملةً في أيِّ مشروعٍ يُركِّز على البيانات بشكلٍ فوري أو مستودعات البيانات النشـطة التكميلية وتوسـيع التخزين التقليدية للبيانات، والانتقال إلى مجـال صُنْع القرارات التشـغيلية والتكتيكية عن طريق تنزيل البيانات بشـكلٍ فوري وتوفير البيانات للمستخدِمين لاتخاذ القرارات الفعَّالة.
- يُعَدُّ أمان البيانات والمعلومات وخصوصيتها من الأمور المهمة بالنسبة للمتخصص في مستودع البيانات.

مصطلحات أساسية:

بطاقة النتائـج المتوازنة (BSC)	نمذجة الأبعاد	الاسـتخراج والتحويـل والتنزيل (ETL)	الخطة التشغيلية
إدارة أداء الأعـمـال (BPM)	DMAIC	سوق البيانات المستقلة	نظم قياس الأداء
تكامل البيانات المكعب	الانتقال لأسفل	مـؤشـر الأداء الرئيسـي (KPI)	مسـتودعات البيانـات الفورية (RDW)
سوق البيانات (DM)	تـكـامـل تـطـبـيـقـات الشركات (EAI)	البيانات الوصفية	ستة سيجما
مـسـتـودع الـبـيـانـات (DW)	مـسـتـودع بـيـانـات المؤسسة (EDW)	OLTP	مخطط مجموعة الثلج

مدير مستودع البيانات (DWA)	تكامل معلومات الشركات (EII)	سوق أوبر	المخطط النجمي
سوق البيانات التابع	جدول الأبعاد	مخزن البيانات التشغيلية (ODS)	

أسئلة للمناقشة:

١- قارن بين تكامل البيانات وETL. وكيف يرتبطان بعضهما ببعض؟

٢- ما هو مستودع البيانات؟ وما هي فوائده؟ ولماذا تُعَدُّ إمكانية الوصول إلى الويب مهمةً في مستودع البيانات؟

٣- هل يُمكِن لـ DM استبدالُ مستودع البيانات أو استكماله؟ قارن وناقش هذه الخيارات.

٤- ناقش العوامل والمزايا الرئيسة لمستودعات البيانات للمستخدِمين النهائيين.

٥- اذكر أوجه الاختلافات و/ أو الشبه بين أدوار مسؤول قاعدة البيانات ومسؤول مستودع البيانات.

٦- اشرح كيف يمكِن أن يؤدي تكامل البيانات إلى مستويات أعلى من جودة البيانات.

٧- قارن بين طريقتَيْ Kimball وInmon نحو تطوير مستودع البيانات. وحدِّد متى يكون كلٌّ منها أكثر فعالية.

٨- ناقش المخاوف الأمنية المتعلقة ببناء مستودع البيانات.

٩- قم بالتحقيق في تنفيذ تطوير مستودع البيانات الحالي من خلال النقل إلى الخارج. واكتب تقريرًا حول هذا الموضوع. ثم ناقش مع زملائك هذه القضية من حيث الفوائد والتكاليف، وكذلك العوامل الاجتماعية.

١٠- يستخدم SAP مصطلح إدارة المؤسسة الإستراتيجية (SEM)، ويستخدم Cognos مصطلح إدارة أداء الشركات (CPM)، ويستخدم Hyperion مصطلح إدارة أداء الأعمال (BPM). هل يشيرون إلى نفس الأفكار الأساسية؟ قدِّم أدلةً لدعم إجابتك.

١١- يشمل BPM خمس عمليات أساسية: وَضْع الإستراتيجية، التخطيط، المراقبة، الفعل، والضبط. اختر إحدى هذه العمليات، وناقش أنواع أدوات البرامج والتطبيقات المتوفرة لدعمها. يُقدِّم الشكل ٣-١٠ بعض التلميحات. ويمكنك أيضًا الاستعانة بقائمة الأدوات المساعدة الخاصة بشركة Bain & Company (bain.com /management_tools /home.asp).

١٢- اختر شركةً عامةً تهمك. باستخدام التقرير السنوي للشركة لعام ٢٠١٦م، قم بإنشاء ثلاثة أهداف مالية إستراتيجية لعام ٢٠١٧م. لكل هدف، حَدِّد هدفًا أو هدفًا إستراتيجيًا. يجب أن تكون الأهداف متسقةً مع الأداء المالي للشركة لعام ٢٠١٦م.

١٣- قارن بين إدارة الأداء وقياس الأداء.

١٤- قم بإنشاء إستراتيجية لشركة افتراضية؛ وذلك باستخدام وجهات النظر الأربع من BSC. مع التعبير عن الإستراتيجية كسلسلة من الأهداف الإستراتيجية. ثم قم بوَضْع خريطة إستراتيجية تصوِّر الروابط بين الأهداف.

١٥- قارن نموذج DMAIC مع عمليات الحلقة المغلقة لـ BPM.

١٦- اختر شركتين تعرفهما جيدًا. ما هي المصطلحات التي يستخدمونها لوَصْف مبادراتهم الخاصة بـ BPM وتركيبات البرامج؟ ثم قارن وميِّز بين عروضهم من حيث تطبيقات ووظائف BPM.

تمارين: جامعة تيراداتا، والتدريبات العملية الأخرى:

١- بالنظر في حالة تطوير مستودع بيانات لشركة Coca-Cola Japan والمتاح على موقع DSS Resources على الويب، http://dssresources.com/cases/coca-colajapan. اقرأ الحالةجيدًا، وأجِب عن الأسئلة التسعة لمزيدٍ من التحليل والمناقشة.

٢- اقرأ مقال الكرة (٢٠٠٥)، وقم بترتيب المعايير (وفقًا لمنظمة حقيقية). في تقرير، اشرح مدى أهمية كلِّ معيار ولماذا؟

٣- اشرح متى يجبُ أن تقوم بتطبيق بنية ثنائية أو ثلاثية الطبقات عند التفكير في تطوير مستودع بيانات.

٤- اقرأ حالة Continental Airlines الكاملة (قصة نجاح تخزين بيانات شائعة للغاية) على شبكة Teradata university.com، وأجب عن الأسئلة.

٥- في teradatauniversitynetwork.com، اقرأ وأجب عن الأسئلة المتعلقة بحالة "أرباح Harrah العالية من معلومات العميل". ثم قم بربط نتائج Harrah بما ستستخدمه شركات الطيران والكازينوهات الأخرى في بيانات العملاء.

٦- في teradatauniversitynetwork.com، اقرأ الإجابة عن أسئلة المهمة "فشل مستودعات البيانات". ونظرًا لأن ثماني حالات موصوفة في هذه المهمة؛ فيمكن تقسيم الفصل إلى ثماني مجموعات، مع تعيين حالة واحدة لكلِّ مجموعة. إضافةً إلى ذلك؛ اقرأ Ariyachandra and

Watson (2006a)، وحَدِّد لكلِّ حالة كيف حدث الفشلُ على أنه لا يرتبط بالتركيز على واحدٍ أو أكثر من عوامل نجاح المرجع.

٧- في teradatauniversitynetwork.com، اقرأ وأجب عن الأسـئلة الخاصة بمهمة «تقنية -Ad Vent: اســتخدام نموذج التحليــل التحليلي لمبيعــات MicroStrategy». يمكن الوصول إلى برنامــج MicroStrategy من موقع TUN. وقد ترغبُ أيضًا في اســتخدام العرض التقديمي لـــ Barbara Wixom حول برنامــج MicroStrategy «العروض التقديميــة للبرنامج النصِّي MicroStrategy Tutorial»، والذي يتوفر أيضًا على موقع TUN.

٨- في teradatauniversitynetwork.com، شاهد ندوات الويب التي تحمل عنوان: «مستودعات البيانات الفورية: الجيل التالي من إدارة بيانات دعم القرار» و«بناء المؤسســة بشكل فوري». ثم اقرأ المقالة التي بعنوان: «البنية المرجعية للمؤسســات بشــكلٍ فوري في تيراداتا: مخطط لمســتقبل تقنية المعلومات»، والتي تتوفر في هــذا الموقع. وقم بوصف كيفية عمل المفاهيم والتقنيات الفورية، وكيف يمكِن اســتخدامها لتوسيع نطاق مستودعات البيانات الحالي وأبنية ذكاء الأعــمال لدعم القــرارات اليومية. واكتــب تقريرًا يوضِّح كيــف أن RDW تقدِّم مَيْزةً تنافســيةً للمؤسســات على وجه التحديد. ثم قم بوصف الصعوبات في مثل هذه التطبيقات والعمليات بالتفصيل، ثم وصف كيفية معالجتها عمليًّا.

٩- في teradatauniversitynetwork.com، شاهد ندوات الويب بعنوان: «النهوض بتكامل البيانات: برامج التشغيل الجديدة والمقاربة الناشئة» و«البحث عن نسخة واحدة للحقيقة: إستراتيجيات لتدعيم الصوامع التحليلية» و«دَمْج البيانات: استخدام ETL وEAI وEII أدوات لإنشاء مؤسسة متكاملــة». واقرأ أيضًا التقرير البحثي «دَمْج البيانــات». قارن بين العروض التقديمية. ما هي القضية الأكثر أهميةً التي تمَّ وصفها في هذه الحلقات الدراسية؟ وما هي أفضل طريقة للتعامل مع الإســتراتيجيات والتحَدِّيات الخاصة بتوحيــد إدارة البيانات وجداول البيانات في بنية مُوحَّدة لمستودعات البيانات؟ قم بإجراء بحثٍ على الويب؛ للتعرُّف على آخر التطوُّرات في هذا المجال. ثم قارن العرض التقديمي بالمادة في النص والمادة الجديدة التي عثرت عليها.

١٠- بالنظر في مستقبل مستودعات البيانات. قم بإجراء بحث على الويب حول هذا الموضوع. أيضًا، اقرأ هاتين المادتين: (Agosta, L. (2006, March 31. مستودعات البيانات في عالم مسطح: اتجاهات لعــام ٢٠٠٦. DM Direct Newsletter; and Geiger, J. G. (2005, November). CIFe: تتطور مع العصر. مراجعة DM، ٣٨-٤١. قارن وقارن نتائجك.

١١- بالوصول إلى teradatauniversitynetwork.com. حَدِّد أحدث المقالات والتقارير البحثية وحالات مستودعات البيانات. مع وَصْف التطورات الأخيرة في هذا المجال. اذكر في تقريرك؛ كيف يتمُّ استخدامُ مستودعات البيانات في ذكاء الأعمال وDSS.

١٢- انتقل إلى موقع YouTube.com وابحث عن «حالات BSI Teradata»؛ إذ تشير BSI إلى "Business Solutions Inc.". حَدِّد ثلاث حالات تخزين بيانات مثيرة للاهتمام، وشاهدها بعناية، واكتب تقريرًا لمناقشة النتائج التي توصَّلت إليها حول مشكلات العمل واقتراحها الحلول الاستقصائية.

١٣- انتقل إلى teradatauniversitynetwork.com. حَدِّد نوع المحتوى «المقالات». تصفح قائمة المقالات، وحَدِّد موقعًا بعنوان: «إدارة أداء الأعمال/ الشركات: تغيير منظر البائع وأهداف السوق الجديدة». بناءً على المقالة، أجب عن الأسئلة التالية:

١- ما هو التركيز الأساسي للمادة؟

٢- ما هي «الوجبات السريعة» الرئيسة من المادة؟

٣- في المقالة، ما هي الوظيفة أو الدور التنظيمي الأكثر مشاركةً بشكلٍ وثيق في CPM؟

٤- ما التطبيقات التي تغطيها CPM؟

٥- كيف تشبه هذه التطبيقات أو تختلف عن التطبيقات التي تغطيها CPM لشركة Gartner؟

٦- ما هي GRC؟ وما هي صِلتها بأداء الشركة؟

٧- ما هي بعض عمليات الاستحواذ الرئيسة التي حدثت في سوق الاجتماع التحضيري للمؤتمر خلال العامين الماضيين؟

٨- حَدِّد اثنتين من الشركات التي تمَّت مناقشتها في المقالة (وليس SAP أو Oracle أو IBM). ما هي إستراتيجيات CPM لكلٍّ من الشركات؟ كيف يفكر المؤلفون فيما يتعلق بهذه الإستراتيجيات؟

١٤- انتقل إلى teradatauniversitynetwork.com. حَدِّد نوع المحتوى «دراسات الحالة». تصفح قائمة الحالات، وحَدِّد موقع «لوحات بيانات فورية في Western Digital». وبناءً على هذه المقالة، أجب عن الأسئلة التالية:

١- ما هو VIS؟

٢- كيف تتشابه أو تختلف بنية VIS عن بنية BPM؟

٣- ما هي أوجه التشابه والاختلاف بين عمليات الحلقة المغلقة من BPM والعمليات في دورة قرار OODA (رَصْد، توجيه، اتخاذ قرار، الفعل)؟

٤- ما هي أنواع لوحات المعلومات الموجودة في النظام؟ هل هي تشـغيلية أم تكتيكية؟ أم هي في الواقع بطاقات أداء؟ اشرح إجابتك.

٥- ما هي المزايا الأساسية التي يوفرها نظام VIS، ولوحات المعلومات في Western Digital؟

٦- ما أنواع الاستشارات التي يمكنك تقديمها إلى شركة تستعدُّ لإنشاء لوحات VIS الخاصة بها؟

مهامُّ الفريق، ومشروعات لعب الأدوار:

١- كانت Kathryn Avery DBA مع سلسـلة متاجر التجزئة على مستوى البلاد (سلسلة كبيرة) على مدى السـنوات السـت الماضيـة. وقد طُلب منهـا في الآونة الأخيرة قيـادة تطوير أول مسـتودع للبيانات في سلسـلة Big Chain. كان المشروع برعاية الإدارة العليا ورئيس قسـم المعلومات. يتمثل الأساس المنطقي لتطوير مستودع البيانات في النهوض بأنظمة التقارير - لا سيما في المبيعات والتسويق وعلى المدى الطويل - لتحسين إدارة علاقات العملاء في السلسلة الكبـيرة. وقد ذهبت Kathryn إلى مؤتمر معهد مسـتودعات البيانـات وكانت تقوم ببعض القـراءة؛ لكنها لا تزال مغمورة بشـأن منهجيات التطوير. إنها تعلـم أن هناك مجموعتين - EDW (Inmon) وDMs (Kimball) الهيـكلي - واللتان تمتلكان مَيْزات قوية بنفس القدر. في البدايـة؛ كانت تعتقد أن المنهجيتين مختلفتان تمامًا؛ لكن نظرًا لأنها درسـتهما بعناية أكثر، فلم تكن على يقين من ذلك. لدى Kathryn عددٌ من الأسئلة التي ترغب في الإجابة عنها:

١- ما هي الفروق الحقيقية بين المنهجيات؟

٢- ما هي العوامل المهمة في اختيار منهجية معينة؟

٣- ماذا ينبغي أن تكون خطوتها التالية في التفكير في المنهجية؟

سـاعد Kathryn في الإجابـة عـن هـذه الأسـئلة. (تـمَّ تعديـل هـذا التمريـن مـن (Duncan, K., Reeves, L., & Griffin, J. 2003, Fall). منظـور ذكاء الأعـمال الخـبراء. Business Intelligence Journal, 8(4), 14-19.

٢- يُعَـدُّ Jeet Kumar مديـر مسـتودعات البيانـات في بنـك إقليمـي كبـير. تـم تعيينـه قبل ٥ سـنوات لتنفيذ مسـتودع بيانات لدعم إسـتراتيجية عمل CRM الخاصة بالبنك. باستخدام مستودع البيانات؛ نجح البنك في دَمْج معلومات العملاء، وفهم ربحية العملاء، وجذب العملاء، وتعزيز علاقات العملاء، والاحتفاظ بالعملاء. على مَرِّ السنين، اقترب مستودع بيانات البنك من الفوريـة من خلال الانتقال إلى عمليات التحديث المتكـررة لمخزن البيانات. والآن؛ يريد البنك

تنفيذ تطبيقات الخدمة الذاتية للعملاء ومراكز الاتصال التي تتطلب بيانات أكثر عذوبةً مما هو متوفر حاليًا في المســتودع. يريد Jeet بعض الدعم في النظر في إمكانيات تقديم بيانات أعذب. ويُعَدُّ أحد البدائل لذلك هو الالتزام الكامل بتنفيذ RDW. لدى مورد ETL استعدادٌ لمساعدته على إجراء هذا التغيير. ومع ذلك؛ فقد تمَّ إبلاغ Jeet عن تقنيات EAI وEII ويتســاءل كيف يمكن أن تتناسب مع خططه. وعلى وجه الخصوص يتساءل Jeet الأسئلة التالية:

١- ما هي بالضبط تقنيات EAI وEII؟

٢- كيف يرتبط EAI وEII بـ ETL؟

٣- كيف يرتبط EAI وEII بـ RDW؟

٤- هل تُعَدُّ كلٌّ من EAI وEII ضروريتين لـ RDW، أم مكملتين لها، أم بديلتين عنها؟

ســاعد Jeet في الإجابــة عــن هــذه الأســئلة. (تــم تعديــل هــذا التمريــن مــن Brobst, S., Levy, E., & Muzilla, C. 2005, Spring. تكامــل تطبيقــات المؤسســات وتكامل معلومات الشركات. Journal Intelligence Journal, 10 (2), 27-33).

٣- قم بمقابلة المســؤولين في كليتك أو المســؤولين التنفيذيين في منظمتك؛ لتحديد كيف يمكن أن يساعدهم مستودعات البيانات في عملهم. اكتب اقتراحًا يَصِف نتائجك. مع تضمين تقديرات التكلفة والفوائد في تقريرك.

٤- راجع قائمة مخاطر مستودعات البيانات الموضَّحة في هذا الفصل، وابحث عن مثالين لكلٍّ منها في الواقع العملي.

٥- يمكنــك الوصول إلى teradata.com، وقراءة الأوراق البيضاء «قياس عائد اســتثمار مســتودع البيانات» و«تحقيق عائد الاستثمار: إسقاط وحصاد قيمة الأعمال لمستودع بيانات المؤسسة». أيضًا، شــاهد الدورة المســتندة إلى الويب «عامل عائد الاســتثمار: كيف يتعامل الممارســون الرائدون مع المســائل المعقدة لقياس عائد الاســتثمار DW.» وَصِفْ أهمَّ القضايا الموضَّحة فيها. قارن هذه القضايا بعوامل النجاح الموصوفة في Ariyachandra وWatson (2006a).

٦- اقرأ المقال بقلم Liddell Avery, K., & Watson, H. J. (2004), Fall. تدريب المستخدمين النهائيين لمســتودعات البيانــات. Journal Intelligence Journal, 9 (4), 40-51 (which available at teradatauniversitynetwork.com). فكِّر في الفئات المختلفة للمستخدِمين النهائيــين، قم بوَصْف الصعوبــات التي تواجههم، ومناقشــة فوائد التدريب المناســب لكلِّ مجموعة. اجعل كلَّ عضو من أعضاء المجموعة يتولى أحد الأدوار، وقم بإجراء مناقشة حول

كيفية أن يكون نوعًا مناسبًا من التدريب على مستودعات البيانات مفيدًا لكلٍّ منكما.

٧- يقدِّم كلُّ مورد من موردي BPM/CPM تقريبًا دراسات حالة على مواقع الويب الخاصة به. باعتبارك فريـق عمل، حَدِّد اثنين من هؤلاء الموِّردين (يمكنك الحصول على أسـمائهم من قوائم Gartner أو AMR). ثم قم باختيار حالتين من كلِّ موقع من هذه المواقع. ولكلِّ حالة قم بإجراء ما يلي:

١- تلخيص المشكلة التي كان العميل يحاول معالجتها.

٢- التطبيقات أو الحلول التي تم تنفيذها.

٣- الفوائد التي يتلقاها العميلُ من النظام.

تمارين الإنترنت:

١- ابحث في الإنترنت للعثور على معلومات حول مستودعات البيانات. حَدِّد بعض مجموعات الأخبار التي لها اهتمامٌ بهذا المفهوم. استكشف ABI/INFORM في مكتبتك ومكتبتك الإلكترونية وGoogle للاطلاع على المقالات الحديثة حول الموضوع. ابدأ بـ tdwi.org وtechnologyevaluation.com والموردين الرئيسـيين: teradata.com وsas.com وoracle.com وncr.com. تحقق أيضًا من cio.com وdmreview.com وdssresources.com وdb2mag.com.

٢- اسـتطلع رأي بعـض أدوات ETL وبعض الموردين. ابدأ مـع fairisaac.com وegain.com. أيضًا استشر dmreview.com (والتي تُسمَّى الآن informationbuilders.com).

٣- اتصل ببعض مورِّدي مستودع البيانات، واحصل على معلومات حول منتجاتهم. أعطِ اهتمامًا خاصًّا للمورِّدين الذين يقدِّمـون أدوات لأغراض مُتعدِّدة، مثل: Cognos، وSoftware A & G، وSAS Institute، وOracle. يوفـر بعـض هؤلاء المورِّدين عروضًـا تجريبيةً مجانيةً عبر الإنترنت. قم بتنزيل عرضٍ أو اثنين منها وقم بتجربتهما. اكتب تقريرًا عن تجربتك.

٤- تصفح teradata.com للتطورات وقصص النجاح حول مستودعات البيانات. اكتب تقريرًا عمَّا اكتشفته.

٥- تصفَّـح teradata.com؛ بحثًـا عـن الأوراق البيضـاء والـدورات المسـتندة إلى الويب حول مسـتودعات البيانات. قراءة السابق ومشـاهدة هذا الأخير. (قسِّم الصف حتى يتمَّ تغطية جميع المصادر) اكتب تقريرًا عمَّا اكتشفته.

٦- قم بإيجاد حالات حديثة من تطبيقات مستودعات البيانات الناجحة. انتقل إلى مواقع مورِّدي مسـتودع البيانات وابحث عن الحالات أو قصص النجـاح. اختر واحدة منها، واكتب ملخصًا مختصرًا لتقديمه إلى زملائك.

المراجع:

- Adamson, C. (2009). The star schema handbook: The complete reference to dimensional data warehouse design. Hoboken, NJ: Wiley.

- Agosta, L. (2006, January). The data strategy adviser: The year ahead—Data warehousing trends 2006. DM Review, 16(1).

- Ariyachandra, T., & Watson, H. (2005). Key factors in selecting a data warehouse architecture. Business Intelligence Journal, 10(3).

- Ariyachandra, T., & Watson, H. (2006a, January). Benchmarks for BI and data warehousing success. DM Review, 16(1).

- Ariyachandra, T., & Watson, H. (2006b). Which data warehouse architecture is most successful? Business Intelligence Journal, 11(1).

- Ball, S. K. (2005, November 14). Do you need a data warehouse layer in your business intelligence architecture? informationmanagement.com/infodirect/20050916/1036931-1. html (accessed June 2016).

- Basu, R. (2003, November). Challenges of real-time data warehousing. DM Review. http://www. information-management.com/specialreports/20031111/7684-1.html (accessed September 2016).

- Benander, A., Benander, B., Fadlalla, A., & James, G. (2000, Winter). Data warehouse administration and management. Information Systems Management, 17(1).

- Bonde, A., & Kuckuk, M. (2004, April). Real world business intelligence: The implementation perspective. DM Review, 14(4).

- Breslin, M. (2004, Winter). Data warehousing battle of the giants: Comparing the basics of Kimball and Inmon models. Business Intelligence Journal, 9(1), 6-20.

- Brobst, S., Levy, & Muzilla. (2005, Spring). Enterprise application integration and enterprise information integration. Business Intelligence Journal, 10 (3).

- Brown, M. (2004, May 9-12). 8 characteristics of a successful data warehouse. Proceedings of the 29th Annual SAS Users Group International Conference (SUGI 29). Montreal, Canada.

- Colbert, J. (2009). Performance management in turbulent times. BeyeNETWORK. http://www.b-eye-network.com/view/10717, (accessed September 2016).

- Davison, D. (2003, November 14). Top 10 risks of offshore outsourcing. Stamford, CT META Group research report, now Gartner, Inc.

- Dull, T. (2015). Data lake vs data warehouse: Key differences. KDnuggets.com. http://www.kdnuggets.com/2015/09/datalake-vs-data-warehouse-key-differences.html (accessed August 2016).

- Eckerson, W. (2003, Fall). The evolution of ETL. Business Intelligence Journal, 8(4).

- Eckerson, W. (2005, April 1). Data warehouse builders advocate for different architectures. Application Development Trends. https://adtmag.com/articles/2005/04/01/data-warehouse-builders-advocate-for-different-architectures.aspx (accessed September 2016).

- Eckerson, W. (2009, January). Performance management strategies: How to create and deploy effective metrics. TDWI Best Practices Report (accessed January 2016). https://tdwi.org/research/2009/01/bpr-1q-performance-management-strategies.aspx (accessed August 2016).

- Eckerson, W., Hackathorn, R., McGivern, M., Twogood, C., & Watson, G. (2009). Data warehousing appliances. Business Intelligence Journal, 14(1), 40-48.

- Edwards, M. (2003, Fall). 2003 Best Practices Awards winners: Innovators in business intelligence and data warehousing. Business Intelligence Journal, 8(4).

- Elson, R., & LeClerc, R. (2005). Security and privacy concerns in the data warehouse environment. Business Intelligence Journal, 10(3).

- Furtado, P. (2009). A survey of parallel and distributed data warehouses. International Journal of Data Warehousing and Mining, 5(2), 57-78.

- Golfarelli, M., & Rizzi, S. (2009). Data warehouse design: Modern principles and methodologies. San Francisco: McGraw-Hill Osborne Media.

- Gupta, P. (2006). Six Sigma business scorecard, 2nd ed. New York: McGraw-Hill Professional.

- Hall, M. (2002, April 15). Seeding for data growth. Computerworld, 36(16).

- Hammergren, T. C., & Simon, A. R. (2009). Data warehousing for dummies, 2nd ed. Hoboken, NJ: Wiley.

- Hatch, D. (2008, January). Operational BI: Getting "real time" about performance. Intelligent Enterprise (accessed March 2016). http://www.intelligententerprise.com/showArticle. jhtml?articleID=205920233 (accessed July 2016).

- Hicks, M. (2001, November 26). Getting pricing just right. eWeek, 18(46).

- Hoffer, J. A., Prescott, M. B., & McFadden, F. R. (2007). Modern database management, 8th ed. Upper Saddle River, NJ: Prentice Hall.

- IBM. (2009). 50 Tb data warehouse benchmark on IBM System Z. Armonk, NY: IBM Redbooks. Imhoff, C. (2001, May). Power up your enterprise portal. E-Business Advise.

- Inmon, W. H. (2005). Building the data warehouse, 4th ed. New York: Wiley.

- Jukic, N., & Lang, C. (2004, Summer). Using offshore resources to develop and support data warehousing applications. Business Intelligence Journal, 9(3).

- Kalido. BP Lubricants achieves BIGS success. kalido.com/collateral/Documents/English-US/CS-BP%20BIGS.pdf (accessed August 2015).

- Kalido. BP Lubricants achieves BIGS, key IT solutions. keyitsolutions.com/asp/rptdetails/report/95/cat/1175 (accessed August 2015).

- Kaplan, R., & Norton, D. (1996). The balanced scorecard: Translating strategy into action. Boston, MA: Harvard University Press.

- Kaplan, R. S., & Norton, D. P. (2005). The balanced scorecard: measures that drive performance. Harvard business review, 83(7), 172.

- Karacsony, K. (2006, January). ETL is a symptom of the problem, not the solution. DM Review, 16(1).

- Kay, R. (2005, September 19). EII. Computerworld, 39(38).

- Kelly, C. (2001, June 14). Calculating data warehousing ROI. SearchSQLServer.com.

- Manglik, A., & Mehra, V. (2005, Winter). Extending enterprise BI capabilities: New patterns for data integration." Business Intelligence Journal, 10(1).

- Matney, D. (2003, Spring). End-user support strategy. Business Intelligence Journal, 8(3).

- Mehra, V. (2005, Summer). Building a metadata-driven enterprise: A holistic approach. Business Intelligence Journal, 10(3).

- Moseley, M. (2009). Eliminating data warehouse pressures with master data services and SOA. Business Intelligence Journal, 14(2), 33-43.

- Murtaza, A. (1998, Fall). A framework for developing enterprise data warehouses. Information Systems Management, 15(4). Nash, K. S. (2002, July). Chemical reaction. Baseline. Issue 8, 27-36.

- Orovic, V. (2003, June). To do & not to do. eAI Journal. pp. 37-43.

- Parzinger, M. J., & Frolick, M. N. (2001, July). Creating competitive advantage through data warehousing. Information Strategy, 17(4).

- Romero, O., & Abelló, A. (2009). A survey of multidimensional modeling methodologies. International Journal of Data Warehousing and Mining, 5(2), 1-24.

- Rosenberg, A. (2006, Quarter 1). Improving query performance in data warehouses. Business Intelligence Journal, 11(1).

- Russom, P. (2009). Next generation data warehouse platforms. TDWI best practices report.tdwi.org/research/reportseries/reports.aspx?pid=842 (accessed January 2016).

- Sen, A. (2004, April). Metadata management: Past, present and future. Decision Support Systems, 37(1).

- Sen, A., & Sinha, P. (2005). A comparison of data warehousing methodologies. Communications of the ACM, 48(3).

- Simons, R. (2002). Performance measurement and control systems for implementing strategy. Upper Saddle River, NJ: Prentice Hall.

- Slywotzky, A. J., & Weber, K. (2007). The Upside: The 7 strategies for turning big threats into growth breakthroughs. Crown Business.

- Solomon, M. (2005, Winter). Ensuring a successful data warehouse initiative. Information Systems Management Journal. 22(1), 26-36.

- Songini, M. L. (2004, February 2). ETL quickstudy. Computerworld, 38(5).

- Thornton, M. (2002, March 18). What about security? The most common, but unwarranted, objection to hosted data warehouses. DM Review, 12(3), 30-43.

- Thornton, M., & Lampa, M. (2002). Hosted data warehouse. Journal of Data Warehousing, 7(2), 27-34.

- Van den Hoven, J. (1998). Data marts: Plan big, build small. Information Systems Management, 15(1).

- Watson, H. J. (2002). Recent developments in data warehousing. Communications of the ACM, 8(1).

- Watson, H. J., Goodhue, D. L., & Wixom, B. H. (2002). The benefits of data warehousing: Why some organizations realize exceptional payoffs. Information & Management, 39.

- Watson, H., Gerard, J., Gonzalez, L., Haywood, M., & Fenton, D.(1999). Data warehouse failures: Case studies and findings. Journal of Data Warehousing, 4(1).

- Zhao, X. (2005, October 7). Meta data management maturity model. DM Direct Newsletter.

الفصل الرابع
التحليلات التنبؤية (١)
عملية، وطرق، وخوارزميات التنقيب في البيانات

أهداف التعلم:

- تعريف التنقيب في البيانات كتقنية تمكينية لتحليلات الأعمال.
- فهم أهداف وفوائد التنقيب في البيانات.
- الدراية الكاملة بمجموعة واسعة من تطبيقات التنقيب في البيانات.
- تعلُّم عمليات التنقيب في البيانات القياسية.
- تعلُّم أساليب وخوارزميات مختلفة من التنقيب في البيانات.
- بناء الوعي بأدوات البرامج الحالية للتنقيب في البيانات.
- فهم قضايا الخصوصية ومشكلات وخرافات التنقيب في البيانات.

بصفةٍ عامة؛ يُعَدُّ التنقيب في البيانات طريقةً لتطوير المعلومات الاستخبارية (أيْ: المعلومات أو المعرفة القابلة للتطبيق) من البيانات التي تُعنى المنظمة بجمعها وتنظيمها وتخزينها. وتستخدمُ المنظمات مجموعةً واسعةً من تقنيات التنقيب في البيانات للحصول على فهمٍ أفضل لعملائها ولعملياتها ولحلِّ المشكلات التنظيمية المعقدة. وندرس في هذا الفصل عملية التنقيب في البيانات كتقنيةٍ تمكينية لتحليلات الأعمال والتحليلات التنبؤية، والتعرُّف على العمليات القياسية لإجراء مشاريع التنقيب في البيانات، وفهم وبناء الخبرات في استخدام تقنيات التنقيب في البيانات الرئيسة، وتطوير الوعي بأدوات البرامج الموجودة، واستكشاف قضايا الخصوصية والخرافات الشائعة والمشكلات التي ترتبط غالبًا بالتنقيب في البيانات.

٤-١ مقال افتتاحي: قسم شرطة ميامي ديد يستخدمُ التحليلات التنبؤية؛ للتنبؤ بالجريمة ومكافحتها:

أصبحت التحليلات التنبؤية والتنقيب في البيانات جزءًا لا يتجزأ من العديد من وكالات إنفاذ القانون، بما في ذلك إدارة شرطة ميامي ديد، والتي لا تقتصر مهمتها على حماية أمن أكبر مقاطعة

في فلوريــدا التي تضــم ٢,٥ مليون مواطن (مما يجعلها ســابع أكبر ولايــة في الولايات المتحدة)، ولكن أيضًا لتوفير مناخٍ آمنٍ وجذاب لملايين السياح الذين يأتون من جميع أنحاء العالم للاستمتاع بالجمال الطبيعي للمقاطعة، وكذلك المناخ الدافئ، والشــواطئ الخلابة. ومع إنفاق الســياح ما يقــرب من ٢٠ مليــار دولار أمريكي كلَّ عام وضخٍّ ما يقرب من ثلث ضرائب المبيعات في فلوريدا؛ فمن الصعب المبالغة في أهمية الســياحة في اقتصاد المنطقة. وعلى الرغم من أن قلةً من ضباط الشرطــة في المقاطعة من المرجَّــح أن يدرجوا التنمية الاقتصادية في وَصْــف وظائفهم؛ فإن كلهم تقريبًا يدركون العلاقة الوثيقة بين الشوارع الآمنة والازدهار الذي تحرِّكه السياحة في المنطقة.

وتُعَدُّ هذه القضية ذات أهميةٍ قصوى بالنســبة للمــلازم أرنولد بالمر، الذي يشرف حاليًّا على قســم التحقيقات في السرقة، والمشرف الســابق على تفاصيل التدخل في قضايا السرقة في الإدارة. إذ يركِّز هذا الفريق المتخصِّص من المحققين بشــكلٍ مُكثَّف على مراقبة النقاط الســاخنة للسرقة بالمقاطعــة وأســوأ المخالفين. ويحتل هو وفريقــه مكاتب متواضعة في الطابــق الثاني من مبنى خرساني حديث المظهر، متفرِّع من شارعٍ مُحاط بالنخيل على الطرف الغربي من ميامي. وقد شهد بالمر الكثيرَ من التغييرات في الســنوات العشر التي قضاها في الوحدة، من أصل ٢٣ عامًا في القوة. ولا يقتصر الأمرُ على ممارســات الشرطة، مثل الطريقة التي اســتخدمها فريقه في تحديد المواقع الساخنة لجرائم الشوارع باستخدام دبابيس ملوَّنة على خريطة.

تقليص خدمات الشرطة:

وشــهد بالمر وفريقه أيضًا تأثير تزايد عدد السكان، وتحوُّل التركيبات السكانية، والاقتصاد المتغير في الشوارع المخولون بحراستها. ومثل أيِّ قوة شرطة جيدة؛ قاموا بتعديل أساليبهم وممارساتهم باستمرار؛ لمواجهــة تحَــدِّي الشرطة الذي اتســع نطاقه وازداد تعقيــده. ولكن على غِرار جميــع فروع حكومة المقاطعــة تقريبًا؛ فــإن ضغوط الموازنة المتزايدة وضعت الإدارة في حالــةِ ضغط بين الطلبات المتزايدة وتقلُّــص الموارد. يقول بالمر، الذي يرى المحققين كمقاتلــين في خط المواجهة ضد موجة متصاعدة من جرائم الشوارع والاحتمالات المرتقبة لموارد متشددة: «كان التحَدِّي الأساسي الذي واجهناه؛ هو كيفية الحَــدِّ من جرائم الشــوارع حتى في الوقت الذي تســببت فيه الموارد المحــدودة في تقليص عددِ رجال الشرطة في الشارع». كان الفريق على مَرِّ السنين مُصرَّحًا له بتجربة أدوات جديدة، كان أبرزها برنامجاً أطلق عليه «الإنفاذ القائم على التحليلات»، والذي استخدم بيانات تاريخ الجريمة كأساس لتعيين فِرَق التحقيق. ويضيف بالمر: «لقد تطوَّرنا كثيراً منذ ذلك الحين في قدرتنا على التنبؤ بالمكان الذي يمكِن أن يحدث فيه سرقة؛ وذلك من خلال كلٍّ من استخدام التحليل وتجربتنا الجماعية الخاصة».

التفكير الجديد في الحالات الباردة:

يتمثل التحَدِّي الأكثر إرباكًا لبالمر وفريقه من المحققين في إغلاق الحالات الأصعب، وهو تحَدٍّ مشتركٍ مع الشرطة في جميع المناطق الحضرية الرئيسة؛ إذ يفتقرُ إلى الأدلة أو الشهود أو الفيديو (أيْ: الحقائق أو الأدلة التي يُمكِن أن تساعد في حَلِّ قضية ما). يقول بالمر: «الأمرُ ليس مفاجئًا؛ لأن الممارسات القياسية التي استخدمناها في توليد العملاء المحتملين، مثل: التحدُّث إلى المخبرين أو المجتمع أو ضباط الدوريات؛ لم تتغيَّر كثيرًا. على كلِّ حال، يعمل هذا النهج بشكلٍ جيد؛ لكن أحد عيوبه أنه يعتمد كثيرًا على خبرات محققينا. فإذا تقاعد هؤلاء المحققون أو انتقلوا إلى أعمالٍ أخرى؛ فإن هذه الخبرات تذهب معهم».

كان لغزُ بالمر هو أن معدل الدوران - بسبب تقاعد العديد من المحققين الأكثر خبرةً لديه - كان في اتجاهٍ تصاعدي. وتوصَّل إلى أنَّ حلَّ هذا اللغز يكمُنُ في ضَخِّ دماء جديدة من محققين أصغر سنًّا، فهم أيضًا يجيدون التعامل مع أنواع المعلومات الجديدة - كرسائل البريد الإلكتروني، ووسائل الإعلام الاجتماعية، وكاميرات المرور، على سبيل المثال لا الحصر - وذلك يمكِّنه أيضًا من الوصول السريع إليها. ولكن كما يقول بالمر: «جاءت المشكلة عندما تحوَّلت مجموعةٌ من المحققين الجُدُد الذين جاءوا إلى الوحدة للبحث عن توجيهات من كبار الضباط. في هذه المرحلة فقط تأكدنا أننا بحاجة إلى طريقةٍ مختلفة لملء فجوة الخبرة في المستقبل».

وقد أدَّت جهودُ بالمر المنصبَّة على التفكير في حَلٍّ لهذا اللغز؛ إلى التوصُّل إلى ما يُسمَّى تكهُّنات السماء الزرقاء. فماذا لو أن المحققين الجدد في الفرقة يمكن أن يطرحوا نفس الأسئلة على قاعدة بيانات الحاسب، كما لو كانوا مخبرًا مخضرمًا؟ زرعت هذه المضاربة في عقل بالمر عددًا لا نهائيًّا من الأفكار.

الصورة الكبيرة تبدأ صغيرة:

أظهر ما كان يتبلور داخل وحدة كشف جرائم السرقة، كيف يمكِن للأفكار الكبيرة أن تأتي من أماكن صغيرة. ولكن الأهم من ذلك؛ هو أنه تبَيَّن أنه لكي تصل هذه الأفكار إلى مرحلة الثبات؛ يجب أن تكون الشروط «الصحيحة» متوافقةً في الوقت المناسب. فعلى مستوى القيادة؛ فإن ذلك يعني أن الشخصية القيادية في المنظمة هي التي تعرف ما يتطلبه الأمر من أجل تغذية الدعم من أعلى إلى أسفل؛ فضلًا عن استقاء المعلومات الذكائية كذلك من أسفل إلى أعلى داخل الهيكل التنظيمي. كان هذا الشخص هو بالمر نفسه. على المستوى التنظيمي؛ شكَّلت وحدة جرائم السرقة نقطة انطلاقٍ جيدة بشكلٍ خاص لنمذجة المبادرة؛ بسبب قيام بعض الجُناة بتكرار

جرائمهم. وفي النهاية؛ فإن قدرة الإدارة على إطلاق العنان للقدرة التحويلية الأوسع نطاقًا لنمذجة المبادرة ستعتمدُ إلى حَدٍّ كبيرٍ على قدرة الفريق على تحقيق النتائج على نطاق أصغر.

عندما أثبتت الاختباراتُ المبكرة والعروض التوضيحية أنها مشجعةٌ - رغم أن النموذج لم يُعطِ نتائج دقيقةً إلا عندما تمَّ تغذيته بتفاصيل الحالات التي تمَّ حلُّها - بدأ الفريق في جذب الانتباه. وقد تلقَّت المبادرة دفعةً قويةً عندما عبَّر مسؤولو وحدة شرطة السرقة عن دعمهم لاتجاه المشروع، وقالوا لبالمر: «إذا استطعتَ القيام بهذا العمل، فافعل». ولكن الأهم من التشجيع - كما يُوضِّح بالمر - هو استعدادهم للدفاع عن المشروع بين الأقسام الأعلى في الإدارة. يقول بالمر: «لم أكن لأترك المشروع على الأرض إذا لم يباركه كبار الضباط؛ لذلك كان دعمهم حاسمًا».

النجاحُ يجلبُ المِصداقية:

بعد تعيين المسؤول الرَّسمي للاتصال بين تقنية المعلومات ووحدة مكافحة السرقة؛ شرَع بالمر في تعزيز أداة نمذجة المبادرة - والتي يُطلَق عليها الآن رسميًا Blue PALMS؛ من أجل برامج النمذجة الرائدة في التحليلات التنبؤية - من خلال تحقيق العديد من النجاحات المتوالية. لم يكن أنصاره من كبار الضباط فحسب؛ بل أيضًا هؤلاء المحققون الذين سيكون دعمهم حاسمًا في تبنِّيه بنجاح كأداة لمكافحة السرقة. وقد قُوبِل بالمر عند محاولته تقديم Blue PALMS بمقاومةٍ عنيفةٍ من قِبل المحققين القدامى، الذين لم يكن لديهم أيُّ مبرر للتخلِّي عن ممارساتهم القديمة. وهنا أدرك بالمر أنَّ الإملاء أو الإكراه لن يجعله ينال وُدَّهم وبالتالي ولا دعمهم. ولذا سيحتاج إلى بناء جسر من المصداقية.

وجد بالمر تلك الفرصة في واحدٍ من أفضل المحققين لديه وأكثرهم خبرةً. في وقتٍ مبكرٍ من التحقيق في السرقة؛ أشار المحقق إلى بالمر أنه كان لديه حدسٌ قوي بمرتكب الجريمة المطلوب ضبطه بالأساس، لاختبار نظام Blue PALMS. وبناءً على طلب المحقق؛ قام محللُ القسم بتغذيةِ النظام بالتفاصيل الرئيسة للجريمة، بما في ذلك طريقة التشغيل. وقامت النماذجُ الإحصائية للنظام بالمقارنة بين هذه التفاصيل وقاعدة البيانات التاريخية للبيانات؛ وذلك بحثًا عن ارتباطات وأوجه تشابهٍ مهمة في تحديد هوية المجرم. وقد تضمَّن التقرير الذي صدر عن العملية قائمةً تضمُّ ٢٠ مشتبهًا تم ترتيبهم حسب ترتيب قوة الجريمة أو احتمال حدوثها. وعندما قام المحلل بتسليم التقرير للمحقق؛ كان المشتبه فيه ضِمن المراكز الخمسة الأولى. وبعد فترةٍ وجيزةٍ من اعتقاله؛ اعترف، واكتسب بالمر أرضًا صلبةً لمشروعه.

على الرغم من أنه كان تمرينًا مفيدًا؛ فقد أدرك بالمر أن الاختبار الحقيقي لم يكن في تأكيد الحدس؛ ولكن في كسر الحالات التي وصلت إلى طريقٍ مسدود. وعلى حَدِّ تعبير بالمر: «كان هذا هو الوضع في سرقة السيارات؛ إذ لا شهود ولا فيديو ولا مسرح جريمة». عندما قام المخبرُ الأقدم في القضية المتوقفة بإجازة بعد ثلاثة أشهر؛ طلب المخبرُ الصغير الذي تمَّ تعيينه له تقرير Blue PALMS. وتم عرضُ صورٍ لأشخاص ذوي حيثية في قائمة المشتبه فيهم، وقام الضحايا بتحديد هوية المشتبه فيه بشكلٍ دقيق؛ مما أدى إلى ختامٍ ناجح للقضية. وكان هذا المشتبه فيه رقم واحد في القائمة.

الحقائق فقط:

لقد تمخَّض النجاحُ المتواصل لـ Blue PALMS عن نجاح بالمر في استقطاب المحققين إلى صَفِّه. ولكن إذا كان هناك جزءٌ من رسالته يَلقى مزيدًا من الصدى لدى المحققين التابعين له؛ فإن الحقيقة هي أنه قد تمَّ مراعاة عدم تغيير أساسيات ممارسات الشرطة عند تصميم Blue PALMS؛ بل أكثر من ذلك فقد تمَّ تحسينها من خلال مَنْحهم فرصةً ثانيةً لتفنيد مزاعمهم تلك بعدم قدرتهم على مجابهة التطورات الحديثة. يقول بالمر: «إنَّ عمل الشرطة هو جوهرُ العلاقات الإنسانية - من حيث التحدُّث مع الشهود، مع الضحايا، مع المجتمع - ونحن لسنا على استعدادٍ لتغيير ذلك؛ بل إن هدفنا هو مَنْح الباحثين إحصاءات واقعية من المعلومات التي لدينا بالفعل والتي قد تُحدِث فرقًا؛ لذا حتى لو نجحنا بنسبة ٥٪ من الوقت؛ فسوف نأخذ الكثير من المذنبين من الشارع».

لقد ساعدت القائمة المتزايدة للحالات الباردة التي حلَّها بالمر في جهوده لتعزيز مزايا Blue PALMS. ولكن في إظهار أين يكمُنُ ولاؤه؛ فهو يرى أن المحققين الذين أغلقوا هذه الحالات الباردة - وليس البرنامج - أنهم يستحقون مزيدًا من تسليط الضوء عليهم، وقد أتى ذلك النَّهْج بثماره كما خطط له بالمر. وبناءً على طلب رئيسه، بدأ بالمر في إجراء اتصالاته كمِنصة للوصول إلى مناطق أخرى في إدارة شرطة ميامي ديد.

شوارع أكثر أمانًا تعني مدينةً أكثر أناقةً:

عندما يتحدث عن تأثير السياحة، وهي مؤشرُ ترابط يمتدُّ من خلال رؤية مدن ميامي ديد الأكثر ذكاءً، ينظر بالمر إلى Blue PALMS على أنها أداةٌ مهمة لحماية أحد أعظم أصول المقاطعة. يقول بالمر: «كان التهديد للسياحة بسبب تزايد الجرائم في الشوارع سببًا رئيسًا في تأسيس الوحدة». ويضيف: «إن حقيقة قدرتنا على استخدام التحليلات والذكاء لمساعدتنا في إغلاق المزيد من القضايا وإخلاء الشوارع من المجرمين، هو خبرٌ سارٌّ لمواطنينا ولصناعة السياحة لدينا».

ماذا يُمكِننا أن نتعلَّم من هذه المقالة القصيرة؟

تتعرَّض هيئات وإدارات تطبيق القانون لضغوطٍ هائلةٍ لتنفيذ مهمتها المتمثلة في حماية الأشخاص ذوي المــوارد المحــدودة. إنَّ البيئة التي يؤدون فيها واجباتهم تصبح أكــثر تحديًّا على نحوٍ متزايدٍ؛ بحيث يتعيَّن عليهم أن يتخذوا باستمرار موقعًا في قلب الحدث وربما يبقون على بُعْد خطوات قليلة لمَنْــع احتمال وقوع كوارث. ويمثل فهم الطبيعة المتغيرة للجريمة والمجرمين تحديًّا مســتمرًا. ففي خِضمِّ هذه التحَدِّيــات؛ يعمل توفُّر البيانات وتقنيات التحليلات؛ من أجل تحليل الأحداث الماضية بشكلٍ أفضل والتنبؤ بالأحداث المستقبلية، ويصبُّ ذلك كله في صالح هذه الوكالات.

وقــد أصبحت البيانات متاحةً أكــثر مما كانت عليه في الماضي. وبالتــالي؛ فإن تطبيق أدوات التحليل المتطوِّرة وأدوات التنقيب في البيانات (أيْ: تقنيات اكتشــاف المعرفة) على هذه المصادر الضخمة والغنية بالبيانات يوفر لهم الرؤية التي يحتاجون إليها لتحضير وتنفيذ واجباتهم بشــكلٍ أفضــل. وبذلك تكون وكالات تطبيق القانون من أهمِّ المســتخدمين للوجــه الجديد للتحليلات؛ إذ يُعَدُّ التنقيب في البيانات مرشــحًا أساســيًّا لفهم أفضل وإدارة أقوى للمهامِّ الحسَّاسة فضلًا عن مستوى عالٍ من الدقة والتوقيت المناسب. وتوضِّح الدراسة المعروضة في المقالة القصيرة الافتتاحية بوضوح قوة التحليلات والتنقيب في البيانات لتوفير رؤيةٍ شاملة لعالم الجريمة والمجرمين؛ من أجل تفاعل وإدارة أفضل وأسرع. وســنعرض في هذا الفصل مجموعةً واســعةً من تطبيقات التنقيب في البيانات التي تقدِّم حلولًا للمشــكلات المعقدة في مجموعة متنوعة من الصناعات والإعدادات التنظيمية؛ إذ يتمُّ اســتخدام البيانات للحصول على رؤيةٍ قابلةٍ للتطبيق بغرض تحســين جاهزية المهمة ورفع الكفاءة التشغيلية وتحقيق الميزة التنافسية.

أسئلة مراجعة على المقالة الافتتاحية:

١- لماذا تتبَنَّى وكالات تطبيق القانون والإدارات، مثل إدارة شرطة ميامي ديد التحليلات المتقدمة والتنقيب في البيانات؟

٢- مــا هي أهمُّ التحَدِّيات التي تواجه وكالات تطبيق القانون والإدارات، مثل إدارة شرطة ميامي ديــد؟ هــل يمكنك التفكير في تحَدِّيات أخرى (غير مذكورة في هذه الحالة) يمكنها الاســتفادة من التنقيب في البيانات؟

٣- ما هي مصادرُ البيانات التي تســتخدمها وكالات تطبيــق القانون والإدارات، مثل إدارة شرطة ميامي ديد لمشاريعها المتعلقة بكلٍّ من النمذجة التنبؤية والتنقيب في البيانات؟

٤- ما نوع التحليلات التي تسـتخدمها وكالات تطبيق القانون والإدارات، مثل إدارة شرطة ميامي ديد لمحاربة الجريمة؟

٥- ماذا تعني «الصورة الكبيرة تبدأ صغيرة» في هذه الحالة؟ اشرح.

٤-٢ مفاهيم وتطبيقات التنقيب في البيانات:

لقـد أصبحت عملية التنقيب في البيانات - وهـي تقنيةٌ جديدةٌ ومثيرةٌ لم تظهر إلا منذ بضع سنوات فقط - عمليةً شائعةً تمارسها أكثرُ المنظمات. وفي مقابلةٍ أجرتها مجلة Computerworld مـع الدكتور Arno Penzias (الحائز على جائزة نوبل وعالم سـابق في مختـبرات Bell) في يناير ١٩٩٩م، قام الدكتور Arno بتحديد البيانات المُستَخرَجة من القواعد التنظيمية باعتبارها تطبيقًا رئيسًـا للشركات في المستقبل القريب. وردًّا على سؤال قديم طرحته جريدة Computerworld حـول «ما هـي التطبيقات القاتلـة في الشركة؟» أجاب الدكتـور Penzias بقولـه: «التنقيبُ في البيانـات». وأردف قائلًا: «سـوف تصبح عمليـة التنقيب في البيانات أكـثر أهميةً، ولن تضطر الـشركات للتخلص من أيِّ شيءٍ يتعلق بعملائهم؛ لأن ذلك سـيكون ذا قيمةٍ كبيرة. وإذا لم تفعل ذلـك فأنت خارج العمل». وبالمثل فقـد ناقش Thomas Davenport (2006) في مقال نشر في Harvard Business Review بأن أحدث سـلاح إستراتيجي للشركات هو اتخاذ القرار التحليلي، وقدَّم أمثلةً لشركات مثل: Amazon.com، وCapital One، وMarriott International، وغيرها من الشركات التي اسـتخدمت التحليلات؛ لفهم عملائها بشـكلٍ أفضل وتحسـين سلاسل التوريد الموسَّـعة لزيادة عوائدهم على الاسـتثمار إلى أقصى حَدٍّ مع توفير أفضـل خدمة للعملاء. حيث يعتمدُ هذا المسـتوى من النجاح بشـكلٍ كبير على فهم الشركة لعملائها، وللموردين، وللعمليات التجارية؛ إضافةً إلى فهمها لسلسلة التوريد الممتدة بشكلٍ جيد.

ومـن الممكـن أن يأتي جزءٌ كبيرٌ من «فهـم العميل» من خلال تحليل كميـة البيانات الهائلة التـي تجمعها الشركة. والجديرُ بالذِّكر أن تكلفة تخزين ومعالجة البيانات قد انخفضت بشـكلٍ كبيرٍ في الماضي القريب. ونتيجةً لذلك؛ فقد نمَت كمية البيانات المُخزَّنة في شـكلٍ إلكتروني بمعدل

Sources: Miami-Dade Police Department: Predictive modeling pinpoints likely suspects based on common crime signatures of previous crimes, IBM Customer Case Studies, www-03.ibm.com/software/businesscasestudies/om/en/corp?synkey=C894638H25952N07; Law Enforcement Analytics: Intelligence-Led and Predictive Policing by Information Builder www.informationbuilders.com/solutions/gov-lea

هائــل. ومع إنشــاء قواعد بيانات كبيرة؛ فقد ظهرت إمكانية تحليــل البيانات المخزَّنة فيها. وقد كان مصطلح التنقيب في البيانات يُستخدَم في الأصل لوَصْف العملية التي يتمُّ من خلالها اكتشافُ أنمــاط غير معروفة مســبقًا في البيانــات. ومنذ ذلك الحين؛ فقد امتدَّ تجــاوز هذا التعريف تلك الحــدود من قِبل بعض مُورِّدي البرامج ليشــمل معظم أشــكال تحليل البيانــات؛ من أجل زيادة المبيعات بشعبية تسمية التنقيب في البيانات. وفي هذا الفصل؛ فإننا نوافق على أن نقبل التعريف الأصلي للتنقيب في البيانات.

وعلى الرغم من أن مصطلح التنقيب في البيانات هو مصطلحٌ جديدٌ نســبيًا؛ فإنَّ الأفكار التي وراءه ليست جديدة؛ إذ تعود جذور العديد من التقنيات المُستَخدَمة في التنقيب في البيانات إلى التحليل الإحصائي التقليدي والذكاء الاصطناعي منذ أوائل الثمانينيات. فلماذا إذًا استحوذت على اهتمام عالم الأعمال؟ فيما يلي بعض الأسباب الأكثر وضوحًا:

- زيادة حِدَّة المنافســة على النطاق العالمي تقودها الاحتياجات المتغيرة باســتمرار للعملاء، وما يريدونه في سوقٍ مُشبَع بشكلٍ متزايد.
- الاعتراف العام بالقيمة غير المستغلة المخفية في مصادر البيانات الكبيرة.
- توحيد وتكامل سِجلات قاعدة البيانات؛ مما يتيح رؤيةً واحدةً للعملاء، والمورِّدين، والمعاملات، وما إلى ذلك.
- توحيد قواعد البيانات ومستودعات البيانات الأخرى في مكانٍ واحدٍ في شكل مستودع البيانات.
- الزيادة المتسارعة في تقنيات معالجة ومستودعات البيانات.
- انخفاض كبير في تكلفة الأجهزة والبرمجيات لمستودعات البيانات ومعالجتها.
- التحرُّك نحو عملية الإزالة (تحويل موارد المعلومات إلى شكلٍ غير مادى) من الممارسات التجارية.

وجديــرٌ بالذكر أنَّ البيانات التي يُولِّدها الإنترنــت تتزايدُ بسرعةٍ؛ من حيث الحجم والتعقيد. وقد تمَّ إنشاءُ كميات كبيرة من البيانات الجينومية، والتي تراكمت في جميع أنحاء العالم. وهناك العديــدُ من التخصُّصات التي تنتج كمياتٍ هائلةً من البيانات على أســاسٍ منتظم. ولذلك يقوم الباحثــون في المجال الطبي والمســتحضرات الصيدلانيــة بإنتاج مســتودعات البيانات التي يمكن اســتخدامها بعد ذلك في تطبيقات التنقيب في البيانات؛ لتحديد أفضل الطرق لتشخيص ومعالجة الأمراض بشكلٍ دقيق، واكتشاف عقاقير جديدة ومُحسَّنة.

وعــلى الجانــب التجــاري؛ فإن أكثر القطاعــات التي تســتخدم تقنية التنقيــب في البيانات، تتمثــل في قطاعات التمويل، والبيــع بالتجزئة، والرعاية الصحية؛ إذ تُســتخدَم تقنية التنقيب في

البيانات؛ للكشــف عن الأنشــطة الاحتيالية والحَدِّ منها، لاســيما في مطالبات التأمين واســتخدام بطاقــات الائتمان (Chan وآخــرون، ١٩٩٩)؛ من أجل تحديد أنمــاط شراء العملاء (Hoffman, 1999)؛ واسترداد العملاء المربحين (Hoffman, 1998)؛ وأيضًا لتحديد قواعد التداول من البيانات التاريخية، والمســاعدة في زيادة الربحية باســتخدام تحليل سلة الســوق. ويُستخدَم التنقيب في البيانات بالفعل على نطاقٍ واسعٍ؛ لتحسين التعامل مع العملاء، ومع تطوُّر التجارة الإلكترونية على نطاق واســع؛ فإن هذا الأمر لا يمكِن أن يصبح أكثر إلحاحًا مع الوقت. انظر الحالة العملية ٤-١؛ للحصول على معلومات حول كيفية استخدام شركة Infinity P&C للتحليلات التنبؤية والتنقيب في البيانات؛ لتحسين خدمة العملاء ومواجهة الاحتيال وزيادة الرِّبح.

حالة عملية ٤-١

Visa هي تعزيزٌ لتجربة العملاء مع الحَدِّ من الاحتيال؛ بالتنقيب في البيانات والتحليلات التنبؤية

عندما بدأت الشركات المُصدِرة للبطاقات لأول مرة باســتخدام برامج قواعد العمل الآلي؛ لمواجهة الاحتيال في البطاقات المصرفية وبطاقات الائتمان - كانت القيودُ المفروضة عــلى هذه التقنية واضحةً تمامًا: اشــتكى العملاءُ من عــدم قدرتهم على الدفع لرحلات عطلاتهــم أو رحــلات العمل المهمة. وتعمل Visa على تحســين تجربــة عملائها؛ من خلال توفير أدوات لتقليل مخاطر الاحتيال؛ إضافةً إلى توفير الخدمات الاستشارية، التي جعلت إستراتيجياتها أكثر فعاليةً. وتهدف Visa من خلال هذا النَّهج إلى تحسين تجربة العملاء، إضافةً إلى تقليل المعاملات غير الصالحة.

وجديــرٌ بالذِّكر أنَّ الشــبكةَ العالميــة للشركة تربط آلاف المؤسســات المالية بملايين التجــار وحامــلي البطاقــات كلَّ يوم. وقــد كانــت Visa رائدةً في مجــال المدفوعات غير النقدية لأكثر من ٥٠ عامًا. كما كانت تدعمُ المؤسســات المالية باستخدام تحليلات SAS؛ وذلــك للحَــدِّ من الاحتيال دون إزعاج العملاء برفــض الدفع دون مبرر. وعندما تقوم Visa بمعالجة إحــدى المعاملات؛ فإنها تقوم بتحليل ما يصلُ إلى ٥٠٠ متغير فريد بشكلٍ فوري لتقيم مخاطر تلك المعاملة. وتستطيع الشركة باستخدام مجموعات هائلة مــن البيانات، بمــا في ذلك النقاط الجديــدة الاحتيالية العالمية وأنمــاط المعاملات - أن تُحــدِّد بدقــة ما إذا كنتَ تشــتري في باريس، أو أن شــخصاً آخــر سرق بطاقة الائتمان الخاصة بك. وقد ذكر Nathan Falkenborg رئيس حلول الأداء في Visa بشــمال آسيا أن: «مــا يعنيه ذلك هو أنه إذا كنتَ ستســافر على الأرجــح؛ فنحن نعرف ذلك ونخبر مؤسســتك المالية حتى لا يتم رفضك في نقطة البيع». وأضاف: «كما أننا سوف نساعدُ

البنك الذي تتعامل معه في تطوير الإستراتيجيات الصحيحة لاستخدام أدوات التأشير وأنظمة تسجيل الدرجات». وتستطيع Visa أن تقوم بعمل تحليلات البيانات الكبيرة؛ ومن المحتمل ان تستطيع النماذج الحديثة وأنظمة تسجيل الدرجات مَنْع ماقيمته ٢ مليار دولار سنوياً من حجم المدفوعات المزوَّرة.

وتُعَدُّ Visa اسمًا معترفًا به عالميًّا، وتسهِّل Visa تحويل الأموال إلكترونيًّا من خلال المنتجات ذات العلامة التجارية التي يُصدِرها الآلاف من شركاء المؤسسات المالية. وقد قامت الشركة بمعالجة ٦٤,٩ مليار صفقة في عام ٢٠١٤م، وتمَّت عمليات شراء بقيمة ٤,٧ تريليون دولار مع بطاقة فيزا في نفس العام.

كما أنَّ Visa لديها القدرة على معالجة ٥٦٠٠٠ رسالة معاملة في الثانية الواحدة، وهذا الرَّقم أكبر من أربعة أضعاف معدل الذروة الفعلي للصفقة حتى الآن. ولا تقوم Visa بالمعالجة والحساب فقط؛ بل هي تَستخدِم باستمرار التحليلات لتبادل الرؤى الإستراتيجية والتشغيلية مع مؤسساتها المالية الشريكة ومساعدتها في تحسين الأداء. ويتمُّ دَعْم هدف هذا العمل من خلال نظام إدارة بيانات قوي. وإضافةً إلى ما سبق؛ فإن فيزا تساعدُ عملاءها على تحسين الأداء من خلال تطوير وتقديم رؤيةٍ تحليلية عميقة.

يقول Falkenborg: «إننا نفهم أنماط السلوك من خلال إجراء التجميع والتجزئة على مستوى دقيق، ونحن نقدِّم هذه الرؤية لشركائنا في المؤسسات المالية، إنها طريقةٌ فعَّالة لمساعدة عملائنا على التواصل بشكلٍ أفضل وتعميق فهمهم للعميل».

وكمثالٍ على الدَّعم التسويقي؛ فقد ساعدت Visa العملاء على الصعيد العالمي في تحديد شرائح العملاء التي يجب أن يُقدَّم لها مُنتَج Visa مختلف. يقول Falkenborg: «يُعَدُّ فهم دورة حياة العملاء أمرًا مهمًّا للغاية، وتوفر فيزا معلومات للعملاء الذين يساعدونهم في اتخاذ الإجراءات وتقديم المنتج المناسب للعميل المناسب قبل أن يصبح عرض القيمة متقادمًا».

كيف يمكِن لاستخدام التحليلات التي في الذاكرة أن تُحدِث فرقًا؟

تَستخدِم Visa مؤخرًا حلاً عاليَ الأداء من SAS، يعتمدُ على الحوسبة في الذاكرة لتشغيل الخوارزمية الإحصائية، وتعلُّم الآلة، ثم تقديم المعلومات بشكلٍ مرئي؛ إذ تقلِّل التحليلاتُ في الذاكرة من الحاجة إلى نقل البيانات وتنفيذ المزيدِ من عمليات تكرار النموذج؛ مما يجعلها أسرع وأكثر دقةً.

ويَصِـف Falkenborg الحـلَّ على أنه مثـل حِفظ المعلومات، بـدلًا من الاضطرار للذهـاب إلى خزانـة الملفات لاسـتردادها. فيقول: «التحليـلات في الذاكرة تأخذ فقط عقلك وتجعله أكبر. كلُّ شيءٍ يمكن الوصول إليه على الفور».

وفي نهاية المطاف؛ فإن التحليلات القوية تسـاعد الشركة على القيام بأكثر من مجرد عملية المدفوعات. يقول Falkenborg: «يُمكِننا تعميق محادثة العميل وخدمة عملائنا بشكلٍ أفضل؛ من خلال مجموعة البيانات الضخمة والخبرات الكبيرة في مجال استخراج بيانات المعاملات، نحن نَستخدِم قدراتنا في الاستشارات والتحليلات لمساعدة عملائنا في مواجهة تحَدِّيات الأعمال وحماية نظام الدفع. وهذا ما نفعله بتحليلات عالية الأداء».

وأضـاف Falkenborg موضحًـا: «إنَّ التحَـدِّي الـذي نواجهـه لا يختلف عن أيِّ تحَدٍّ يواجه أيَّ شركةٍ تديرُ وتسـتخدم مجموعة البيانات الضخمة؛ فهو يتمثل في كيفية اسـتخدامنا لجميع المعلومات اللازمة لوَضْع حَلٍّ للتحَدِّيات التي تواجه الأعمال - سواء كان ذلك من شأنه تحسين نماذج الاحتيال لدينا، أو مساعدة العميل على التواصل بشكلٍ أكثر فعاليةً مع زبائنه». ثم اسـتطرد قائلًا: «تمكِّننا التحليلات في الذاكرة من أن نكون أكـثر فطنةً؛ مع سرعـة معالجة ١٠٠ نظام تحليلي؛ حيث يُمكِـن لبياناتنا ولعلماء القرار لدينا التكرار بشكلٍ أسرع».

وتسـمحُ التحليلات التنبؤية السريعة والدقيقة لشركـة Visa بتقديم خدمةٍ أفضل للعملاء؛ إضافةً إلى خدمات استشـارية مُتخصِّصة؛ مما يساعدهم على النجاح في صناعة المدفوعات المتغيرة سريعًا بشكل يومي.

أسئلة للمناقشة:

١- ما هي التحَدِّيات التي واجهتها Visa، وباقي صُنَّاع بطاقات الائتمان؟

٢- كيف تمكَّنَت Visa من تحسين خدمة العملاء مع الاحتفاظ بتحسين الاحتيال؟

٣- ما هي التحليلات في الذاكرة، ولماذا تُعَد ضروريةً؟

Source: "Enhancing the customer experience while reducing fraud (SAS® Analytics) - High-performance analytics empowers Visa to enhance customer experience while reducing debit and credit card fraud."

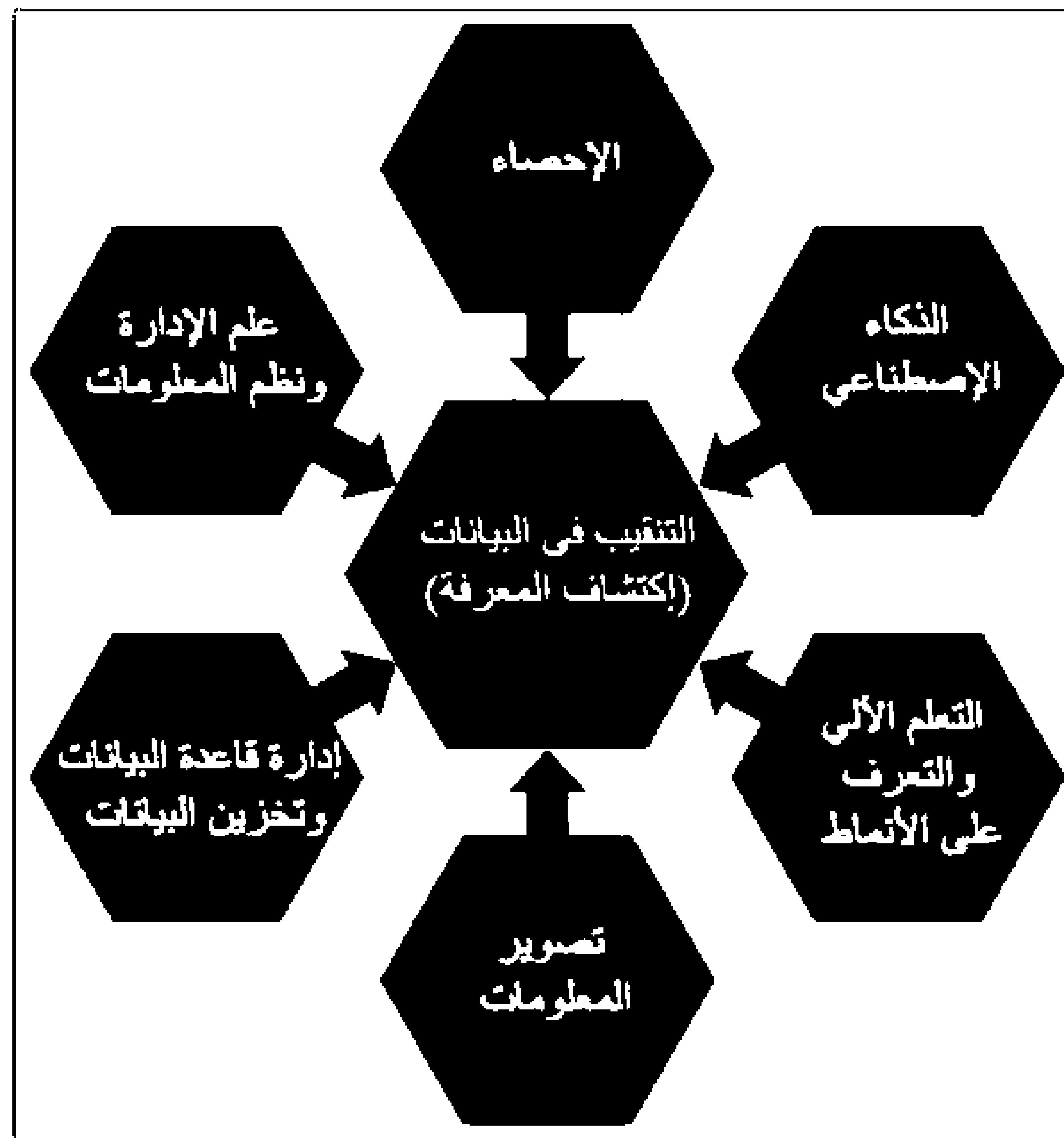
التعريفات، والخصائص، والفوائد:

إنَّ مصطلح التنقيب في البيانات ببساطة؛ هو مصطلحٌ يُستَخدَم لوَصْف اكتشاف أو «استخراج» المعرفة من كميات كبيرة من البيانات. وبالقياس؛ فإنه عند النظر، يستطيع المرء أن يدرك بسهولة أنَّ مصطلح التنقيب في البيانات ليس صحيحًا؛ إذ إنه يُطلَق على استخراج الذهب من الصخور أو التراب، ويُشار إليه باسم استخراج «الذهب» بدلًا من الاستخراج «الصخري» أو «الترابي». ولذلك؛ فقد كان الأصحُّ أن يُقال: «استخراج المعرفة» أو «اكتشاف المعرفة» بدلًا من التنقيب في البيانات. وعلى الرغم من عدم التوافق بين المصطلح ومعناه؛ فقد أصبح التنقيبُ في البيانات هو اختيار الأغلبية. فالعديدُ من الأسماء الأخرى المرتبطة بالتنقيب في البيانات تشمل: استخراج المعرفة، وتحليل النمط، وعلم الآثار، وحصاد المعلومات، والبحث عن الأنماط، وتجريف البيانات.

ومن الناحية الفنية؛ فإن عملية التنقيب في البيانات هي عمليةٌ تَستَخدِم تقنيات إحصائية، ورياضية، وأساليب الذكاء الاصطناعي لاستخلاص وتحديد المعلومات المفيدة والمعرفة (أو الأنماط) الفرعية من مجموعات كبيرة من البيانات، ومن الممكن أن تكون هذه الأنماط في شكلِ قواعد العمل أو الانتماءات أو الارتباطات أو الاتجاهات أو نماذج التنبؤ (انظر: Nemati and Barko, 2001). ومعظم الأدبيات تُعرِّف التنقيبَ في البيانات بأنه: «عمليةٌ غير بديهية لتحديد أنماط صالحة وجديدة ومفيدة ومفهومة في نهاية المطاف في البيانات المخزَّنة في قواعد البيانات المنظمة»؛ إذ يتمُّ تنظيمُ البيانات في سجلات منظمة من قِبل المتغيرات الفئوية والترتيبية والمستمرة (Fayyad وآخرون، ١٩٩٦ ص ٤٠-٤١). وفي هذا التعريف؛ نجد أن معاني المصطلح الأساسي كما يلي:

- تشيرُ العملية إلى أن التنقيب في البيانات يضمُّ العديد من الخطوات المتكرِّرة.
- غير بديهي؛ بمعنى أن هناك بعض عمليات البحث أو الاستنتاج من نوع التجارب؛ بمعنى أنه ليس بسيطًا مثل حساب الكميات المُحدَّدة مسبقًا.
- صالح؛ بمعنى أن الأنماط المكتشفة يجب أن تنطبق على البيانات الجديدة بدرجة كافية من اليقين.
- غير مألوفة؛ بمعنى أنَّ الانماط لم تكن معروفةً من قبل للمُستخدِم في سياق النظام الجاري تحليله.
- الفائدة المحتمَلة؛ وتعني أنَّ الأنماط المُكتشَفة يجب أن تؤدي إلى فائدةٍ بعيدة للمستخدِم أو المهمة.
- وأخيرًا؛ مفهومًا؛ بمعنى أنَّ النمط يجب أن يكون منطقيًّا من الناحية التجارية إلى الدرجة التي تجعل المُستخدِم يقول: «امممم يبدو الأمر معقولًا؛ لماذا لم أفكر في ذلك، وإن لم يكن هذا على الفور، فعلى الأقل بعد تحليل بعض نتائج العمل».

إنَّ التنقيـبَ في البيانــات ليس نظامًا جديدًا؛ بــل هو تعريفٌ جديدٌ لاســتخدام العديد من التخصُّصات؛ إذ يتمُّ وَضْع بيانات الاســتخراج بشــكلٍ مُحكَم عند تلاقي العديد من التخصصات، والتي تشمل الإحصاء والذكاء الاصطناعي وتعلم الآلة وعلوم الإدارة ونظم المعلومات (IS) وقواعد البيانات (انظر الشــكل ٤-١)؛ إذ إنَّ عملية التنقيب في البيانات تســعى جاهدةً إلى التحســين في اســتخراج المعلومات والمعارف المفيدة من قواعد البيانات الكبيرة، عن طريق استخدام التقدُّم في جميع التخصُّصات الســابقة. وعلى الرغم من أن هذا المجال لا يزال في بداياته؛ فإنه جذب إليه الكثيرَ من الاهتمام في وقتٍ قصير للغاية.



شكل ٤-١: التنقيب في البيانات هو مزيجٌ من التخصصات المتعددة

وفيما يلي الخصائصُ والأهداف الرئيسية للتنقيب في البيانات:

- غالبًا ما يتمُّ دفن البيانات في أعماق قواعد البيانات الكبيرة جدًّا، والتي تحتوي أحيانًا على بيانات من عِدَّة ســنوات. وفي العديدِ من الحالات؛ فإن تنظيفَ البيانات وتوحيدها يتمُّ في مســتودع البيانات؛ إذ يُمكِن تقديمُ البيانات بتنسيقات متنوعة (انظر: الفصل ٢ لتصنيف موجز للبيانات).
- بيئة التنقيب في البيانات؛ هي عادةً بنية العميل/ الخادم، أو بنية IS على الويب.

- تســاعد الأدوات الجديدة المتطوِّرة، والتي تشــمل أدوات التصويــر المتقدِّمة، على إزالة مصادر المعلومات المدفونة في ملفات الشركة أو السِّــجلات العامة الأرشيفية. والبحث عن ذلك يشمل تدفــق ومزامنــة البيانات؛ للحصول على النتائــج الصحيحة. كما يستكشــف العمالُ المتقنون للتنقيــب في البيانات أيضًا جــدوى البيانات الضعيفة (أيْ: النص غــير المُنظَّم المُخزَّن في أماكن مثل قواعد بيانات Lotus Notes أو الملفات النصية على الإنترنت أو الشــبكات الداخلية على مستوى المؤسسة).
- غالبًا ما يكون عاملُ التنقيب في البيانات مُستَخدَمًا نهائيًّا، ويتمُّ تمكينه بواسطة مثاقيب البيانات وأدوات الاســتعلام القوية الأخرى لطرح أســئلةٍ مُخصَّصة والحصول على إجابات بسرعة؛ وذلك من خلال استخدام مهارة برمجة محدودة أو منعدمة.
- وكثــيرًا مــا ينطوي التنقيبُ الموسَّــع في البيانات عــلى إيجاد نتيجة غير متوقعــة، ويتطلب من المستخدِمين النهائيين التفكيرَ بصورة إبداعية خلال العملية، بما في ذلك تفسير النتائج.
- يتمُّ الجَمْع بين أدوات التنقيب في البيانات بسهولةٍ مع جداول البيانات وأدوات تطوير البرمجيات الأخرى. وبالتالي؛ يُمكِن تحليلُ البيانات المتحصَّل عليها، ونَشْرها بسرعة وسهولة.
- نظــرًا للكميات الكبيرة من البيانات وجهود البحث المكثفة؛ فإنه من الضروري في بعض الأحيان استخدام المعالجة المتوازية للتنقيب في البيانات.

ويُمكِن للشركة التي تســتفيدُ بشكلٍ فعَّال من أدوات وتقنيات التنقيب في البيانات؛ الحصولُ على مَيْزة تنافســية إســتراتيجية والحفاظ عليهــا؛ إذ يُوفِّر التنقيبُ في البيانــات للمنظمات بيئةً مُعزِّزة لاتخاذ القرارات لاستغلال الفرص الجديدة عن طريق تحويل البيانات إلى سلاحٍ إستراتيجي. ولتفاصيل أكثر حول فوائد التنقيب في البيانات انظر: Nemati and Barko (2001).

كيف يعملُ التنقيب في البيانات؟

من خلال اســتخدام البيانات الحالية ذات الصِّلة والتي تــمَّ الحصولُ عليها من داخل وخارج المنظمة؛ يســتطيع التنقيبُ في البيانات أن يبني نماذج لاكتشــاف الأنماط بين السِّمات المعروضة في مجموعــة البيانات. وهــذه النماذجُ هي التمثيلات الرياضيــة (العلاقات/ الارتباطات الخطية البســيطة أو العلاقات المعقَّدة وغير الخطية) التي تحَدِّد الأنماط بين سِــمات الأشياء (كالعملاء أو الأحداث) الموضَّحة في مجموعة البيانات. وبعض هذه الأنماط تفسيرية (تشرح العلاقات المتبادلة والصِّلات بين السِّمات)، في حين أن البعض الآخر تنبوئية (تُنبِّئ بالقِيَم المستقبلية لبعض السِّمات).

وبشكلٍ عام؛ فإن التنقيبَ في البيانات يسعى إلى تحديدِ أربعة أنواعٍ رئيسية فقط من الأنماط:

١- تحديد الترابط في مجموعات مشـتركة من الأشياء، مثل: البيرة والحفاضات تسير معًا في تحليل سلة السوق.

٢- التنبؤات تُوضِّح طبيعة الأحداث المستقبلية لأحداث معينة بناءً على ما حدث في الماضي، مثل: التنبؤ بالفائز في Super Bowl، أو التنبؤ بالحرارة المطلقة في يومٍ معين.

٣- التجميـع ويعملُ على تحديدِ المجموعات الطبيعية للأشـياء اسـتنادًا إلى خصائصها المعروفة، مثل: تعيين العملاء في شرائح مختلفة بناءً على التركيبة السكانية وسلوكيات الشراء السابقة.

٤- العلاقات التسلسـلية، وتعمل على كَشْـف علاقـات الأحداث الزمنية، مثـل التنبؤ بأن عميلًا مصرفيًّا موجودًا لديه بالفعل حسـاب فحص سيفتح حساب توفير متبوعًا بحساب استثمار في غضون عام.

وقد تمَّ اسـتخراجُ هذه الأنواع من الأنماط يدويًّا من البيانات من قِبل البشر على مدى قرون، غير أن تزايد حجم البيانات في العصر الحديث جعل هناك حاجة إلى طرق أكثر تلقائية. ومع ازدياد حجم مجموعات البيانات وتعقيدها؛ ازداد التحليل المباشر للبيانات اليدوية بشكلٍ كبيرٍ باستخدام أدوات معالجـة البيانـات الآلية وغير المباشرة التي تسـتخدم منهجيات وأسـاليب وخوارزميات مُعقدة. ويُشَـار لمظاهر مثل هذا التطور من الوسائل الأوتوماتيكية وشِبه الأوتوماتيكية لمعالجة البيانات الكبيرة باسم التنقيب في البيانات.

وبشكلٍ عام؛ فإنه يمكِن تصنيف مهامِّ التنقيب في البيانات إلى ثلاث فئاتٍ رئيسية، هي: التنبؤ، والارتباط، والتجميع. واستنادًا إلى الطريقة التي يتمُّ بها استخراجُ الأنماط من البيانات التاريخية؛ يُمكِـن تصنيف أنظمة تعلُّـم خوارزميات التنقيب في البيانات إلى التنقيـب في البيانات في وجود إشراف أو بدون إشراف، وتشـتمل بيانات التدريب على كلٍّ من الصفات الوصفية (أيْ: المتغيرات المسـتقلة، أو متغيرات القـرار) إضافةً إلى خاصية الفئـة (أيْ: متغير المُخْـرَج أو متغير النتيجة). وفي المقابـل؛ فمـع التعلُّم بدون إشراف، فإن بيانات التدريب تشـمل فقـط الصفات الوصفية. ويوضح الشـكل ٤-٢ تصنيفًا بسيطًا لمهامِّ التنقيب في البيانات، جنبًا إلى جنب مع أساليب التعلُّم والخوارزميات الشائعة لكلِّ مهمة من مهام التنقيب في البيانات.

نوع التعلم	خوارزميات التنقيب في البيانات	مهام وطرق التنقيب في البيانات
		التنبؤ
خاضعة للإشراف	أشجار القرارات والشبكات العصبية و آلة المتجهات الداعمة وGA، وNaive Bayes، وkNN	التصنيف
خاضعة للإشراف	الإنحدار الخطي وغير الخطي و ANN وأشجار الإنحدار SVM، وkNN، وGA	الإنحدار
خاضعة للإشراف	طرق الإنحدار الذاتي وطرق حساب المعدل والتبسيط الأسي وARIMA	السلاسل الزمنية
		التجميع
غير خاضعة للإشراف	Apriori, OneR ,ZeroR, Eciat,GA	سلة السوق
غير خاضعة للإشراف	تعظيم التوقعات وخوارزمية Apriori والتطبيق المبني على الرسم البياني	تحليل الرابط
غير خاضعة للإشراف	خوارزمية Apriori والنمو FP- والتطبيق المبني على الرسم البياني	تحليل التتابع
		التقسيم إلى قطاعات
غير خاضعة للإشراف	K-means وتعظيم التوقعات (EM)	التكتل
غير خاضعة للإشراف	K-means وتعظيم التوقعات (EM)	تحليل القيم الشاذة

شكل ٤-٢: تصنيف مُبسَّط لمهامٍّ وطرق وخوارزميات التنقيب في البيانات

- **التنبــؤ:** يُشــار إلى التوقع عــادةً على أنه إخبارٌ عن المســتقبل. وهو يختلــف عن التخمين البسيط بعِدَّة أشياء، مثل: مراعاة الخبرات والآراء والمعلومات الأخرى التي لها صِلةٌ بالتنبؤ. وهناك مصطلــحٌ آخر يرتبط عادةً بمصطلح التنبؤ هــو التوقُّع، وعلى الرغم من اعتقاد الكثيرين أن هذين المصطلحين مترادفان؛ فإنَّ هناك اختلافًا دقيقًا بل حاسماً بين المصطلحين. فالتنبؤ يُستخدَم إلى حَدٍّ كبيرٍ في الإشارة إلى الخبرة والرأي، في حين أن التوقع هو البيانات والنموذج القائم. ومن أجل زيادة الثقــة يُمكِن للمرء أن يدرج المصطلحــات التي لها صِلةٌ ببعضها، مثل: التخمــين، والتنبؤ، والتوقع على التوالي. وفي مصطلحات التنقيب في البيانات؛ يتمُّ اســتخدامُ التنبؤ والتوقع بشــكلٍ مترادف، ويُســتخدَم مصطلح التنبؤ باعتباره التمثيل المشــترك للفعل. وتبعًا لطبيعة ما يتمُّ التنبؤ به؛ فإنه

يُمكِن تسـمية التنبؤ بشـكلٍ أكثر تحديدًا على أنه تصنيفٌ (حيث يكون الشيء الذي يتمُّ التنبؤ به، مثل: تنبؤات الغد والتي تُصنِّف حالة الجو إلى «ممطر» أو «مشمس») أو انحدارٌ (حيث يتمُّ التنبؤ بشيء، مثل: درجة حرارة الغد، والتي تكون عبارة عن رقم حقيقي، مثل «٦٥ درجة فهرنهايت»).

- **التصنيـف:** ويُقصَد به التصنيـف أو الإشراف التعريفي، وهو الأكثر شـيوعًا في جميع مهامٍّ التنقيـب في البيانات. والهدف مـن التصنيف؛ هو تحليلُ البيانات التاريخيـة المُخزَّنة في قاعدة البيانات وإنشاء نموذج يُمكِنه التنبؤ بالسلوك المستقبلي تلقائيًّا. ويتكوَّن هذا النموذج المُستحدَث من تعميمات لسجلات مجموعة بيانات التدريب، والتي تساعد على تمييز الفئات المُحدَّدة مسبقًا. ونأمل أن يُسـتخدَم هذا النموذج بعد ذلك للتنبؤ بفئات السـجلات الأخرى غير المصنَّفة، والأهم من ذلك، للتنبؤ بدقة بالأحداث المستقبلية الفعلية. وتشمل أدوات التصنيف الشائعة: الشبكات العصبية وأشـجار القـرار، والتحليل اللوجسـتي وتحليـل التمييز (من الإحصـاءات التقليدية)، والأدوات الناشئة، مثل: المجموعات الخام، وآلات متجه الدعم (SVMS)، والخوارزميات الجينية. وقد تلقت تقنيات التصنيف القائمة على الإحصاء (على سبيل المثال: الانحدار اللوجستي وتحليل التمييـز) نصيبها مـن الانتقادات - وهي أنها تضـع افتراضات غير واقعية حـول البيانات، مثل: الاستقلالية والوَضْع الطبيعي - مما يَحِدُّ من استخدامها في مشاريع التنقيب في البيانات من النوع التصنيفي. وتنطوي الشـبكات العصبية على تطوير هياكل رياضية (تشـبه إلى حَدٍّ ما الشـبكات العصبيـة البيولوجيـة في الدماغ البشري) والتـي لديها القدرة على التعلُّم من التجارب السـابقة المقدمـة في شـكل مجموعات بيانات جيدة التنظيم. فهي تميـلُ إلى أن تكون أكثر فعاليةً عندما يكـون عددُ المتغيرات المعنيَّة كبيرًا إلى حَدٍّ ما والعلاقات فيـما بينها معقدة وغير دقيقة. وجديرٌ بالذِّكر أن الشـبكات العصبية لها عيوبٌ كما أن لها مزايا. فعلى سـبيل المثال: عادةً ما يكون من الصعـب للغاية توفـير مُبرِّرات جيدة للتنبؤات التي تقوم بها الشـبكة العصبيـة. إضافةً إلى أن الشـبكات العصبية تحتـاج إلى تدريبٍ كبير. ولسـوء الحظ؛ فإن الوقت الـلازم للتدريب يزداد بشـكلٍ كبيرٍ مع زيادة حجم البيانات، وبشـكلٍ عام؛ فإنه لا يمكِن تدريب الشبكات العصبية على قواعد بيانات كبيرة جدًّا. وهذه العوامل وغيرها سـاعدت على تحديث إمكانية تطبيق الشبكات العصبية في المجالات الغنية بالبيانات.

وتقوم أشـجارُ القرار بتصنيف البيانات إلى عددٍ مُحدَّد مـن الفئات بناءً على قِيَم متغيرات الإدخال. وأشـجار القرار؛ هي في الأساس تسلسـلٌ هرمي للبيانات ثم تكون أسرع بشـكلٍ ملحوظ من الشبكات العصبيـة. فهـي أكثر ملاءمةً للبيانات الفئوية والفترة الزمنية. ولذلك؛ فإن دَمْج المتغيرات المسـتمرة في إطار شجرة القرارات يتطلب تفكيرًا؛ أيْ: تحويل المتغيرات الرقمية القيِّمة المستمرة إلى النطاقات والفئات.

أمَّا الفِئة ذات الصِّلة من أدوات التصنيف؛ فهي اســتقراءُ القاعدة. وهي على عكس شــجرة القرار، ومع اســتقراء القاعدة؛ فإن العبارات حينئذٍ تنتــجُ عن بيانات التدريب مباشرةً، ولا يجب أن تكــون ذات طبيعــة هرمية. وهنــاك تقنياتٌ أخرى حديثة مثــل SVM، ومجموعات صعبة، وخوارزميات جينية، تَجِد طريقها تدريجيًّا إلى ترسانة من خوارزميات التصنيف.

- **التجميع:** هو عبارة عن تجميع أقســام مجموعةٍ من الأشــياء (على ســبيل المثال: كائنات، أحداث، معروضة في مجموعة بيانات منظمة) إلى شرائح (أو مجموعات طبيعية) يتشارك أعضاؤها في خصائص متشــابهة. فهو على عكس التصنيف؛ إذ إنَّ تجميــع تصنيفات الفئات غير معروف. وإذ إنَّ الخوارزميات تمرُّ عبر مجموعة البيانات؛ فإن تحديد القواســم المشــتركة للأشــياء يتمُّ بناءً على خصائصها، ومن ثمَّ يتمُّ إنشــاءُ المجموعات. ونظرًا لأن المجموعات يتم تحديدها باستخدام خوارزميــة من النــوع التجريبي؛ ولأن الخوارزميات المختلفة قد تنتهــي بمجموعات مختلفة من المجموعات لمجموعة البيانات نفسها، قبل أن يتمَّ وَضْع نتائج تقنيات التجميع للاستخدام الفعلي؛ فقد يكون من الضروري وجود خبير لتفســير، وربما تعديــل، المجموعات المقترحة. وبعد تحديد المجموعات المعقولة، يمكن استخدامها لتصنيف وتفسير البيانات الجديدة.

وليس من المُســتَغرَب أن تشــتمل تقنيات التجميع على تحقيق الدرجة المُثلى. والهدف من التجميع هو إنشاء مجموعات؛ بحيث يكون للأعضاء داخل كلِّ مجموعة أقصى درجة من التشابه والأعضــاء عــبر المجموعات لديهم الحَــدُّ الأدنى من التشــابه. وتتضمَّن تقنيــات التجميع الأكثر استخدامًا وسائل k (من الإحصائيات) وخرائط التنظيم الذاتي (من تعلُّم الآلة)، وهي بنيةٌ فريدةٌ للشبكة العصبية طوَّرها Kohonen (1982).

وغالبًا ما تستخدم الشركاتُ بفاعلية نُظُم التنقيب في البيانات لديها؛ لإجراء التجزئة في السوق مــع التحليل التجميعي. وتحليل المجموعة؛ يُعَدُّ وســيلةً لتحديد فئــات العناصر؛ بحيث تكون العناصرُ الموجودة في مجموعة مشــتركة بعضها مع بعض بشــكلٍ أكبر مــن العناصر الموجودة في مجموعات أخرى. ويُمكِن اســتخدامه في تقســيم العملاء وتوجيه المنتجات التســويقية المناسبة إلى الشرائح في الوقت المناســب وبالشكل المناسب وبالسعر المناسب. يُستَخدَم تحليل المجموعة أيضًا؛ لتحديد التجمُّعات الطبيعية للأحداث أو الكائنات؛ بحيث يُمكِن تحديدُ مجموعةٍ مشــتركة من سمات هذه المجموعات لوصفها.

- **الاقتران:** إنَّ الجمعيات، أو تعلُّم قواعد الترابط في التنقيب في البيانات؛ تقنيةٌ شائعةٌ ومدروسةٌ جيدًا لاكتشــاف العلاقات المثيرة بين المتغــيِّرات في قواعد البيانات الكبــيرة. وبفضل تقنيات جَمْع

البيانات الآلية، مثل: ماسحات الباركود؛ فإن استخدام قواعد الرَّبط لاكتشاف الإجراءات المنتظمة بين المنتجات في المعاملات الكبيرة التي سجَّلتها أنظمة نقاط البيع في محلات السوبر ماركت أصبحت مهمةً في اكتشاف معرفة مشتركة في صناعة البيع بالتجزئة. وفي سياق صناعة البيع بالتجزئة، غالبًا ما يُطلَق على الاستخراج الخاص بقواعد الارتباطات: تحليل السوق. وهناك اثنان من المشتقات الشائعة الاستخدام؛ وهما: رابطة الاستخراج من القاعدة وتحليل الارتباط وتسلسل الاستخراج. وباستخدام تحليل الارتباطات، يتمُّ اكتشاف الارتباط بين العديدِ من الأشياء ذات الأهمية تلقائيًّا، مثل الارتباط بين صفحات الويب والعلاقات المرجعية بين مجموعات مؤلفي النشرات الأكاديمية. ومع الاستخراج التسلسلي؛ يتمُّ فَحْص العلاقات؛ من حيث ترتيب حدوثها لتحديد الجمعيات مع مرور الوقت. وتتضمَّن الخوارزميات المُستَخدَمة في استخراج القاعدة المشتركة Apriori الشهير (حيث يتمُّ التعرُّف على العناصر المتكرِّرة) وFP-Growth، وOneR، وZeroR، وEclat.

التصويرُ والتنبُّؤ بالسلسلة الزمنية؛ هناك تقنيتان ترتبطان في الغالب بالتنقيب في البيانات؛ وهما التصويرُ والتنبؤ بالسُّلسلة الزَّمنية. ويمكن استخدامُ التمثيل المرئي بالاقتران مع تقنيات التنقيب في البيانات الأخرى؛ للحصول على فَهْم أكثر وضوحًا للعلاقات الأساسية. ومع ازديادِ أهمية التصوير في السنوات الأخيرة؛ ظهر مصطلحٌ جديدٌ، وهو التحليلات المرئية. وتكمُنُ الفكرة في الجَمْع بين التحليلات والتصوير في بيئةٍ واحدة لتسهيل إنشاء المعرفة بشكلٍ أسهل وأسرع. وقد تمَّ تغطية التحليلات المرئية بالتفصيل في الفصل ٣. وفي توقع السلسلة الزمنية؛ تتكوَّن البيانات من قِيَم المتغيِّر نفسه الذي يتمُّ التقاطه وتخزينه مع مرور الوقت في فواصل زمنية منتظمة. ثم يتمُّ استخدام هذه البيانات لتطوير نماذج التوقع لاستقراء القِيم المستقبلية للمتغير نفسه.

التنقيب في البيانات مقابل الإحصاءات:

هناك الكثيرُ من القواسم المشتركة بين التنقيب في البيانات والإحصاءات. فكلاهما يبحث عن العلاقات داخل البيانات. ومعظم الناس يتصلون بالإحصائيات «أساس التنقيب في البيانات». والفرق الرئيس بين الاثنين؛ هو أن الإحصائيات تبدأ بفرضية مُحددة بدقة؛ في حين تبدأ عملية التنقيب في البيانات ببيان اكتشاف غير مُحدَّد المعالم. وتجمع الإحصائيات بيانات عينة (أيْ: بيانات أولية) لاختبار الفرضية، في حين أنَّ التنقيب في البيانات والتحليلات تستخدم جميع البيانات الموجودة (أيْ: البيانات الثانوية الملاحظة في الغالب) لاكتشاف أنماط وعلاقات جديدة. وهناك فرقٌ آخر يأتي من حجم البيانات التي يستخدمونها؛ إذ يبحث التنقيب في البيانات عن مجموعات البيانات «الكبيرة» قدر الإمكان؛ في حين تبحث الإحصائيات عن الحجم الصحيح للبيانات (إذا

كانت البيانات أكبر من المطلوب/ المطلوب للتحليل الإحصائي، يتمُّ اســتخدام عينة من البيانات). إنَّ معنــى «البيانات الضخمة» يختلف إلى حَـدٍّ ما بين الإحصاءات والتنقيب في البيانات. وهناك القليــلُ من المئــات إلى الآلاف من نقاط البيانات الكبيرة بما يكفي للتحليل الإحصائي، غير أنَّ عِدَّة ملايين إلى بضع مليارات من نقاط البيانات تُعَدُّ كبيرةً بالنسبة لدراسات التنقيب في البيانات.

حالة عملية ٤-٢

احتفاظ شركة Dell بذكاء وفعالية تحليلاتها في القرن الحادي والعشرين

لقــد غيَّرت الثورة الرقمية كيفية تســوق الناس؛ إذ تشــير الدراســات إلى أنه حتى العملاء التجاريين يســتخدمون المزيدَ من حلولِ البحث عــن رحلة الشراء الخاصة بهم عــبر الإنترنت قبل ارتباطهــم ببائع. ومن أجل التنافس؛ فإن بعض الشركات، مثل: شركة Dell تقــوم بتحويل نماذج المبيعات والتســويق لدعم هــذه المتطلبات الجديدة. ومع ذلــك؛ فإن القيام بهذا الأمر على نحوٍ فعَّال يتطلب حلَّ «البيانات الكبيرة» والذي يُمكِنه تحليل قواعد بيانات الشركات جنبًا إلى جنب مع معلومات غير منظمة من مصادر مثل clickstreams والشبكات الاجتماعية.

لقــد تطــوَّرت Dell لتصبح رائــدةً في مجال التقنية من خلال اســتخدام العمليات الفعَّالــة القائمة عــلى البيانات؛ إذ يُمكِــن للموظفين الحصولُ على نتائــج قابلة للقياس لعقود قادمة من الزمان؛ وذلك من خلال اســتخدام تطبيقات المؤسســات لدعم الرؤية وتســهيل العمليات، مثل: إدارة علاقات العملاء (CRM) والمبيعات والمحاسبة. وعندما أدركــت Dell أن العملاء يقضون وقتًــا أطول في البحث عن المنتجات عبر الإنترنت قبل الاتصال بمندوب مبيعات؛ أرادت تحديث النماذج التســويقية وفقًا لذلك؛ بحيث يمكنها تقديم الأنواع الجديدة من الخدمــات المخصَّصة والدعم الذي يتوقعه العملاء. لإجراء هذه التغييرات، ويحتاج موظفو التســويق إلى المزيدِ من البيانات حول ســلوك العملاء عبر الإنترنت. كما يحتاج الموظفون إلى طريقةٍ أســهل لتضييق الرؤية من خلال العديد مــن أدوات ذكاء الأعمال ومصادر البيانات. وقد ذكــر Drew Miller المدير التنفيذي لتحليلات الســوق والأفــكار في Dell أن: «هنــاك العديدَ من المعلومــات المتاحة عن عادات التســوق عبر الإنترنــت وخارج الإنترنت للعملاء. نحــن بحاجةٍ فقط إلى إعطاء موظفي التسويق حلًا سهل الاستخدام يُمكِنه استيعاب كل ذلك، وتحديد الأنماط وتقديم توصيات حول الإنفاق على التسويق والأنشطة».

إنشاء فريق ذكي لتعزيز عائد الاستثمار (ROI) باستخدام ذكاء الأعمال والتحليلات:

ومن أجل تحسين إستراتيجية ذكاء الأعمال والتحليلات والاتصالات العالمية؛ فقد أنشأت Dell فريقًا مهمًّا لتقنية المعلومات. كما أنشأ المديرون التنفيذيون نموذجًا ذكيًّا للحوكمة في الفريق؛ حتى يتمكَّن من الاستجابة بسرعة لمتطلبات ذكاء الأعمال المتطوِّرة للموظفين ومتطلبات التحليلات وتقديم عائد استثمار سريع. فعلى سبيل المثال: نجد أنه إضافةً إلى امتلاك حرية التعاون مع مجموعات الأعمال الداخلية؛ فإنه يتمُّ تمكين فريقِ العمل من تعديلِ الأعمال وعمليات تقنية المعلومات باستخدام الإستراتيجيات الذكية والمبتكرة. كما يجبُ أن يكرِّس الفريق أكثر من ٥٠٪ من جهوده لتحديد وتنفيذ المكاسب السريعة لذكاء الأعمال ومشاريع التحليلات التي عادةً ما تكون صغيرة جدًا بالنسبة لقائمة أولويات «A» بقسم تقنية المعلومات في Dell. كما يجب على الفريق أيضًا أن ينفق ما لا يقل عن ٣٠٪ من وقته في التبشير داخل مجموعات الأعمال الداخلية لرفع مستوى الوعي حول التحويلات التي تتمتع بها ذكاء الأعمال - فضلًا عن الفرص المتاحة للتعاون.

وقد كان أحدُ المشروعات الأولى لفريق العمل؛ هو إيجاد حلٍّ جديدٍ لبراءات الاختراع والتحليل يُعرف باسم Workbench Analytics. وقد ركَّز تطبيقه المبدئي على مجموعة مختارة من حالات الاستخدام حول ارتباطات العملاء التجاريين عبر الإنترنت وخارجها. وقد تمَّ دَعْم هذا المجهود من قِبل مؤسسات تقنية المعلومات والتسويق في Dell. يقول Fadi Taffal، مدير شركة تقنية المعلومات في شركة Dell: «كانت هناك رغبةٌ في توسيع نطاق استخدام هذا الحل لدعم العديد من أنشطة المبيعات والتسويق في أقرب وقتٍ ممكن. ومع ذلك؛ كنا نعلم أننا نستطيع بناء حلٍّ أكثر فعالية إذا قمنا بتطويره من خلال الجهود السريعة المتسارعة».

سوق معلومات واحدٍ ضخم (One Massive Data Mart) لتسهيل مصدر واحدٍ للحقيقة:

ومن خلال العمل بجدية مع التسويق؛ فإن مهندسو فريق العمل يستخدمون إستراتيجيات تطوير البرمجيات الضعيفة، والعديد من التقنيات لإنشاء سوق بيانات قابلٍ للتوسُّع بشكلٍ كبير؛ إذ يَستخدِم الحلُّ الشامل العديدَ من التقنيات والأدوات لتمكين أنواع مختلفة من مستودعات البيانات، والمعالجة البارعة، وأنشطة التشغيل الآلي. فعلى سبيل المثال: يقوم المهندسون بمستودعات البيانات غير المُنظمة من مصادر الوثائق الرقمية/ الاجتماعية على الخوادم التي تقوم بتشغيل Apache Hadoop؛ إذ يستخدمون صيغة Teradata Aster؛ لدَمْج واستكشاف كميات كبيرة من بيانات العملاء من مصادر

أخرى بشكلٍ فوري. وبالنسبة للعديدِ من متطلبات تحويل البيانات والتشغيل الآلي؛ فإنَّ الحلَّ يتضمَّن استخدام برمجيات Toad من Dell، وتحديدًا Toad Data Point، وToad Intelligence Central، إضافةً إلى Dell Statistica؛ إذ توفر برمجية Toad Data Point واجهةً ملائمةً للأعمال لمعالجة البيانات والتشغيل الآلي، وهي فجوةٌ حَرِجةٌ في النظام البيئي. وبالنسبة للنماذج التحليلية المتقدِّمة؛ فإن النظام يستخدم Dell Statistica، والذي يوفر إعداد البيانات والتحليلات التنبؤية واستخلاص البيانات والتعرُّف الآلي، والإحصاءات، وتحليلات النصوص، والتوقيع والإبلاغ، ونشر النموذج ومراقبته. ويستفيد المهندسون أيضًا من هذا الحل لتطوير نماذج تحليلية يُمكِن من خلالها فحص جميع البيانات المتباينة وتقديم صورة دقيقة لسلوك التسوق لدى العملاء. وتوفر الأدوات اقتراحات لتحسين الخدمة؛ إضافةً إلى مقاييس عائد الاستثمار لإستراتيجيات تعدُّد المنتجات التي تشمل التسويق عبر الويب والمكالمات الهاتفية وزيارات المواقع.

وفي غضون عِدَّة أشهر كان المهندسون يستخدمون منصة التحليلات التسويقية الأولية. كما قامت فرقة العمل بالتخطيط لتوسيع إمكانات الحلِّ حتى تتمكَّن من تحليل البيانات من مصادر أكثر، وتقديم تصورات إضافية، وقياس عوائد أنشطة القنوات الأخرى، مثل: التغريدات، والنصوص، ورسائل البريد الإلكتروني، ومشاركات وسائل الإعلام الاجتماعية.

توفير أكثر من ٢,٥ مليار دولار في التكاليف التشغيلية:

ومن خلال الحلول الجديدة؛ فقد قامت شركة Dell بالفعل بإزالة العديد من تطبيقات ذكاء الأعمال الخاصة بطرف ثالث. وقد ذكر Chaitanya Laxminarayana، مدير برنامج التسويق في شركة Dell أنه: «على الرغم من أننا في المراحل الأولى من طرح منبر التسويق التحليلي؛ فإننا وفَّرنا حوالي ٢,٥ مليون دولار في تكاليف الاستعانة بمصادر خارجية للمورِّدين». «وإضافةً إلى ذلك؛ يحصل الموظفون على إحصاءات أسرع وأكثر تفصيلًا». ونظرًا لأن Dell تقوم بتحجيم نطاق تحليل التسويق؛ فإنها ستعمل على التخلُّص التدريجي من تطبيقات ذكاء الأعمال الأخرى التابعة لجهات خارجية؛ مما يقلل من التكاليف ويُعزِّز الكفاءة.

تسهيل ٥,٣ مليون دولار في الإيرادات:

يتمتَّع موظفو التسويق الآن بالرؤية التي يحتاجون إليها لتحديد الاتجاهات الناشئة في ارتباطات العملاء، وتحديث النماذج وفقًا لذلك. وقد ذكر Laxman Srigiri، مدير

قسـم التحليـلات التسـويقية في شركة Dell: «لقـد حققنا بالفعـل ٥,٣ مليون دولار كإيرادات متزايدة من خلال إطلاق برامج تسـويقية أكثر تخصيصًا وكشف فرص جديدة من خلال قاعدة البيانات الكبيرة لتحليلات التسويق». «إضافةً إلى ذلك؛ لدينا برامج في مسارها لتوسيع نطاق هذا التأثير عِدَّة مرات في السنوات الثلاث المقبلة».

وعلى سـبيل المثال: يُمكِن للموظفين الآن مشاهدة جدول زمني لتفاعلات العميل عبر الإنترنت وبدون اتصال مع Dell، بما في ذلك عمليات الشراء، وصفحات Dell الخاصة بموقع الويب التي زارها العميل، والملفات التي قاموا بتنزيلها. وإضافةً إلى ذلك؛ يتلقى الموظفون اقتراحاتٍ لقواعـد البيانات؛ لمعرفة وقت وكيفية الاتصال بأحد العملاء؛ إضافةً إلى عناوين URL لصفحات محدَّدة ينبغي عليهم قراءتها لمعرفة المزيد عن التقنيات التي يبحث عنها العميل. يقول Srigiri: «كان من الضروري أن نفهم المتطلبات المتغيِّرة؛ حتى نتمكن من الحفـاظ عـلى ذكائنا. والآن بعد أن أصبح لدينا هذه الرؤيـة؛ يمكننا بسرعة تطوير نماذج تسويقية أكثر فاعليةً توفر المعلومات المخصَّصة والدعم الذي يتوقعه العملاء».

أسئلة للمناقشة:

١- ما هو التحَدِّي الذي واجهته Dell، والذي أدَّى إلى رحلة تحليلاتها؟

٢- ما هو الحلُّ الذي قامت شركة Dell بتطويره وتنفيذه؟ وماذا كانت النتائج؟

٣- كشركة التحليلات نفسـها، اسـتخدمت Dell عروض خدماتها لأعمالها الخاصة، هل تعتقد أنه من الأسهل أو الأصعب على الشركة تذوق الدواء الخاص بها؟ وضِّح ذلك.

Source: Dell: Staying agile and effective in the 21st century. Dell Case Study, software.dell.com/casestudy/dell-staying -agile-and- effective-in-the-21st-century881389. Used by permission from Dell.

أسئلة للمراجعة على قسم ٤-٢:

١- عرِّف التنقيب في البيانات. واذكر لماذا يُوجَد العديدُ من الأسماء والتعريفات المختلفة للتنقيب في البيانات؟

٢- ما هي العواملُ الأخيرةُ التي زادت من شعبية التنقيب في البيانات؟

٣- هل يُعَدُّ التنقيبُ في البيانات تخصصاً جديداً؟ وضح ذلك.

٤- اذكر بعض الطرق الرئيسة وخوارزميات التنقيب في البيانات؟

٥- ما هي الاختلافات الأساسية بين المهامِّ الرئيسة للتنقيب في البيانات؟

٤-٣ تطبيقات التنقيب في البيانات:

لقد أصبح التنقيبُ في البيانات أداةً شائعةً في معالجة العديد من المشكلات والفرص التجارية المعقــدة. وقــد ثبُت أنه ناجحٌ للغاية ومفيــدٌ في العديد من المجالات، وبعضهــا مُبيَّن في الأمثلة التمثيلية التالية. والهدف من العديدِ من تطبيقات التنقيب في البيانات التجارية؛ هو حلُّ مشكلة مُلحَّة، أو استكشاف فرصة عمل ناشئة لخلق مَيْزة تنافسية مُستدَامة.

- **إدارة علاقات العملاء:** إدارة علاقات العملاء (CRM) هي امتدادُ التسويق التقليدي. ويتمثل الهدف من إدارة علاقات العملاء CRM في إنشــاء علاقات فردية مع العملاء من خلال تطوير الفهــم الدقيــق لاحتياجاتهم ورغباتهم. ومــع قيام الشركات ببناء علاقات مــع عملائها بمرور الوقت من خلال مجموعةٍ متنوعةٍ من التفاعلات (مثل: استفسارات المنتج والمبيعات، وطلبات الخدمــة، ومكالمــات الضمان، ومراجعات المنتجــات، واتصالات وســائل التواصل الاجتماعي)؛ فإنهــا تجمع كمياتٍ هائلةً من البيانات. وعند الجمع بين السِّــمات الديموغرافية والاجتماعية والاقتصادية، يمكِن استخدام هذه البيانات الزاخرة بالمعلومات من أجل:

١- تحديد المستجيبين/ المستفيدين المحتملين للمنتجات/ الخدمات الجديدة (كتصنيف العملاء).

٢- فهم الأسباب الجذرية لاستنزاف العملاء؛ لتحسين استبقاء العملاء (أيْ: تحليل التغيير).

٣- اكتشــاف ارتباطات متغيِّر الوقت بين المنتجات والخدمات؛ لتحقيق أقصى قدرٍ من المبيعات والقيمة للعملاء.

٤- تحديد العملاء الأكثر ربحيةً واحتياجاتهم التفضيلية؛ لتعزيز العلاقات وزيادة المبيعات.

- **الخدمات المصرفية:** يستطيع التنقيبُ في البيانات أن يساعدَ البنوك على القيام بما يلي:

١- أتمتة عملية طلب القرض من خلال التنبؤ الدقيق للمستفيدين الأكثر احتمالاً للتعثر.

٢- كشف بطاقات الائتمان المزوَّرة والمعاملات المصرفية عبر الإنترنت.

٣- تحديد سُــبُل تعظيم قيمة العميل عن طريق بَيْع المنتجات والخدمات التي من المرجَّح أن يقوموا بشرائها.

٤- تحسين العائد النقدي؛ من خلال التنبؤ بدقة بالتدفق النقدي على الكيانات المصرفية (مثل: أجهزة الصراف الآلي، الفروع المصرفية).

- **البَيْع بالتجزئة والنقل والإمداد:** في صناعة البيع بالتجزئة؛ يمكن استخدام التنقيب في البيانات في:

١- التنبؤ بحجم مبيعات دقيقة في مواقع البيع بالتجزئة المحَدَّدة لتحديد مستويات المخزون الصحيح.

٢- تحديد علاقات المبيعات بين المنتجات المختلفة (مع تحليل سلة السوق)؛ من أجل تحسين تخطيط المتجر وتحسين ترويج المبيعات.

٣- مستويات الاستهلاك المتوقعة من أنواع مختلفة من المنتجات (على أساس الظروف الموسمية والبيئية) لتحسين النقل والإمداد، وبالتالي زيادة المبيعات.

٤- اكتشاف أنماط مثيرة للاهتمام في حركة المنتجات (خاصةً بالنسبة للمنتجات التي لها فترة صلاحية محدودة؛ لأنها عُرْضة لانقضاء الصلاحية وقابلية الاستخدام والتلوث) في سلسلةِ التوريد من خلال تحليل بيانات التعرُّف على الترددات اللاسلكية والحسية (RFID).

- **التصنيع والإنتاج:** يُمكِن للمصنِّعين استخدامُ التنقيب في البيانات لكلٍّ من:

١- التنبؤ بفشل الآلات قبل حدوثها؛ من خلال استخدام البيانات الحسية (تمكين ما يُسمَّى الصيانة المستندة إلى الشروط).

٢- تحديد الشذوذ والقواسم المشتركة في نظم الإنتاج؛ لتحسين القدرة التصنيعية.

٣- اكتشاف أنماط جديدة؛ لتحديد وتحسين جودة المنتج.

- **السَّمسرة وتداول الأوراق المالية:** يَستخدم الوسطاءُ والتجارُ التنقيب في البيانات؛ من أجل:

١- التنبؤ بوقت ومقدار تغيُّر أسعار السَّندات.

٢- التنبؤ بنِطاق واتجاه تقلُّبات الأسهم.

٣- تقييم تأثير قضايا وأحداث مُعيَّنة على حركات السوق ككل.

٤- تحديد ومَنْع الأنشطة الاحتيالية في تداول الأوراق المالية.

- **التأمين:** تَستخدِم صناعة التأمين تقنيات التنقيب في البيانات؛ من أجل:

١- توقُّع مبالغ المطالبة الخاصة بتكاليف المِلكية والتغطية الطبية؛ لتحسين خطط الأعمال.

٢- تحديد خطط الأسعار المُثلى على أساس تحليل المطالبات وبيانات العملاء.

٣- التنبؤ بالعملاء الأكثر احتمالًا لشراء سياسات جديدة ذات سِمات خاصة.

٤- تحديد ومَنْع مدفوعات المطالبة غير الصحيحة والأنشطة الاحتيالية.

- **أجهزة الحاسب والبرمجيات:** يُمكِن استخدامُ التنقيب في البيانات في:

١- توقع حالات فشل مُحرِّك الأقراص قبل حدوثها فعليًّا.

٢- تحديد وتصفية محتوى الويب غير المرغوب فيه ورسائل البريد الإلكتروني.

٣- اكتشاف ومَنْع ثغرات أمن شبكات الحاسب.

٤- تحديد منتجات البرمجيات غير الآمنة.

- **الحكومة والدفاع:** يحتوي التنقيبُ في البيانات أيضًا على عددٍ من التطبيقات العسكرية. حيث يمكن استخدامه في:

١- توقع تكاليف نقل الأفراد والمعدَّات العسكرية.

٢- التنبؤ بحركات الخَصْم، وبالتالي تطوير إستراتيجيات أكثر نجاحًا للارتباطات العسكرية.

٣- التنبؤ باستهلاك الموارد؛ لتحسين التخطيط والميزانية.

٤- تحديد فئات من التجارب والإستراتيجيات والدروس المستفادة من العمليات العسكرية لتبادل المعرفة بشكلٍ أفضل في جميع أنحاء المنظمة.

- **صناعة السفر (شركات الطيران، الفنادق/ المنتجعات، شركات تأجير السيارات):** يُستَخدَم التنقيب في البيانات في العديدِ من مجالات صناعة السفر؛ إذ يتمُّ استخدامها بنجاح في:

١- التنبؤ بمبيعات الخدمات المختلفة (أنواع المقاعد في الطائرات، أنواع الغرف في الفنادق/ المنتجعات، أنواع السيارات في شركات تأجير السيارات)؛ من أجل تسعير الخدمات على النحو الأمثل لزيادة الإيرادات إلى أقصى حَدٍّ كدالة للمعاملات المتغيرة مع الوقت (يُشار إليها عادةً باسم إدارة العائد).

٢- توقُّع الطلب في مواقع مختلفة؛ من أجل تخصيص موارد تنظيمية محدودة بشكلٍ أفضل.

٣- تحديد العملاء الأكثر ربحيةً وتزويدهم بالخدمات الشخصية؛ للحفاظ على أعمالهم المتكررة.

٤- الاحتفاظ بالموظفين القيِّمين من خلال تحديد الأسباب الجذرية التي تؤدي إلى تناقص العملاء والتصرُّف بناءً على ذلك.

- **الرعاية الصحية:** يحتوي التنقيبُ في البيانات على عددٍ من تطبيقات الرعاية الصحية؛ إذ يمكن استخدامه في:

١- تحديد الأشخاص الذين ليس لديهم تأمينٌ صحي والعوامل الكامنة وراء هذه الظاهرة غير المرغوبة.

٢- تحديد علاقات التكلفة/ الفائدة الجديدة بين المعالجات المختلفة لتطوير إستراتيجيات أكثر فعالية.

٣- توقُّع مستوى ووقت الطلب في مواقع الخدمة المختلفة؛ لتخصيص الموارد التنظيمية على النحو الأمثل.

٤- فهم الأسباب الكامنة وراء تناقص العملاء والموظفين.

- **الدواء:** يجب النظرُ إلى استخدام التنقيب في البيانات في الطب باعتباره مكملًا قيمًا للبحوث الطبية التقليدية، والتي هي أساسًا سريرية وبيولوجية في الطبيعة؛ إذ يمكن لتحليلات التنقيب في البيانات:

١- تحديد أنماط جديدة لتحسين قابلية بقاء المرضى المصابين بالسرطان.

٢- التنبؤ بمعدلات نجاح مرضى زرع الأعضاء لتطوير سياسات مطابقة أفضل للأعضاء.

٣- تحديد وظائف الجينات المختلفة في الكروموسوم البشري (المعروف باسم علم الجينات).

٤- اكتشاف العلاقات بين الأعراض والأمراض (وكذلك الأمراض والعلاجات الناجحة) لمساعدة المهنيين الطبيين في اتخاذ قرارات مستنيرة وصحيحة في الوقت المناسب.

- **صناعة أو مجال التسلية والترفيه:** تَستخدِم صناعة الترفيه بنجاح عملية التنقيب في البيانات في:

١- تحليل بيانات المشاهد؛ لتحديد البرامج التي يتمُّ عرضها خلال وقت الذروة، وكيفية زيادة العائد من خلال معرفة مكان إدراج الإعلانات.

٢- توقع النجاح المالي للأفلام قبل أن يتمَّ إنتاجها؛ لاتخاذ قرارات استثمارية وتحسين العائدات.

٣- التنبؤ بالطلب في أماكن مختلفة وأوقات مختلفة؛ من أجل تحديد مواعيد أفضل للأحداث الترفيهية وتخصيص الموارد على النحو الأمثل.

٤- وَضْع سياسات تسعير مثالية لزيادة الإيرادات.

- **الأمن الداخلي، وتنفيذ القانون:** يحتوي التنقيب في البيانات على عددٍ من تطبيقات أمن الأراضي وتطبيق القانون. وغالبًا ما يتم استخدام التنقيب في البيانات في:

١- تحديد أنماط السلوكيات الإرهابية، (وللحصول على مثالٍ على استخدام التنقيب في البيانات لتتبُّع تمويل أنشطة الإرهابيين، انظر: الحالة العملية ٤-٣).

٢- اكتشاف أنماط الجريمة (مثل: المواقع، والتوقيت، والسلوك الإجرامي، والسِّمات الأخرى ذات الصِّلة)؛ للمساعدة في حَلِّ القضايا الجنائية في الوقت المناسب.

٣- التنبؤ بالهجمات البيولوجية والكيميائية المحتملة على البنية التحتية الحيوية للأمة والقضاء عليها عن طريق تحليل البيانات الحِسِّية ذات الأغراض الخاصة.

٤- تحديد وإيقاف الهجمات الضارة على الهياكل الأساسية الحيوية للمعلومات (غالبًا ما تُسمَّى حرب المعلومات).

- **الرياضات:** يُستخدَم التنقيبُ في البيانات لتحسين أداء فِرق الرابطة الوطنية لكرة السلة (NBA) في الولايات المتحدة؛ إذ تستخدم فِرق البيسبول الرئيسية في الدوري التحليلات التنبؤية والتنقيب في البيانات؛ للاستفادة على النحو الأمثل من مواردها المحدودة لموسم الفوز (انظر: مقالة Moneyball في الفصل الأول). وفي الواقع؛ فإن معظم - إن

لم يكن كل - الرياضات المحترفة في الوقت الحاضر توظف خبراء البيانات كما تستخدم التنقيب في البيانــات لزيادة فرصهم في الفــوز. ولا تقتصر تطبيقات التنقيب في البيانات على الرياضات الاحترافيــة. ففي مقالة عام ٢٠١٢م؛ قام كلٌّ من Delen وCogdell و(2012) Kasap بتطوير نمــاذج التنقيب في البيانات للتنبؤ بنتائج الرابطة الرياضية الجامعية الوطنية (NCAA) ونتائج لعبة الكرة باستخدام مجموعة واسعة من المتغيرات حول إحصائيات الألعاب السابقة للفريقين المتعارضــين (مزيــدٌ من التفاصيل حول هذا يتمُّ توفير دراســة الحالة في الفصــل الثاني). وقد اســتخدم (2012) Wright مجموعــةً متنوعةً مــن المتنبئين لفحص دوري بطولة كرة السَّــلة للرجال في الرابطة الوطنية لأمريكا الشمالية (a.k.a. March Madness).

حالة عملية ٤-٣

التحليل التنبؤي والتنقيب في البيانات يساعدُ في وقف تمويل الإرهاب

أكَّــد الهجومُ الإرهابي على مركز التجارة العالمي في ١١ ســبتمبر ٢٠٠١م، على أهمية الذكاء المفتوح المصدر. وقد أعلن قانون الوطنية الأمريكية وإنشاء وزارة الأمن الداخلي الأمريكية عن احتماليةٍ لتطبيق تقنية المعلومات وتقنيات التنقيب في البيانات؛ للكشف عن غســيل الأموال والأشــكال الأخرى لتمويل الإرهاب. وتركِّــز وكالات إنفاذ القانون على أنشــطة غسل الأموال عن طريق المعاملات العادية من خلال المصارف وغيرها من منظمات الخدمات المالية. وتركِّز وكالات إنفاذ القانون الآن على تسعير التجارة الدولية كأداة لتمويل الإرهاب. وقد استخدم تجار غسيل الأموال التجارة الدولية؛ لنقل الأموال بصمــت خارج البلاد دون جــذب انتباه الحكومة. يتمُّ تحقيق هذا التحويل عن طريق المبالغــة في تقدير الواردات وانخفاض قيمة الصادرات. فمثلًا يمكن للمســتورد المحلي والمُصَدِّر الأجنبي أن يشــكلا شراكةً ويتجاوزا قيمة الواردات، وبالتالي تحويل الأموال من البلد الأم؛ مما يــؤدي إلى جرائم تتعلق بالاحتيال الجمركي، والتهرُّب من ضريبة الدخل، وغسيل الأموال. ويمكن أن يكون المُصدِّر الأجنبي عضوًا في منظمة إرهابية.

وتركِّز تقنيات التنقيب في البيانات على تحليل البيانات المتعلِّقة بمعاملات الاســتيراد والتصديــر من وزارة التجارة الأمريكية والكيانات المرتبطة بالتجارة. ويتم تتبُّع أســعار الواردات التي تتجاوز الحَدَّ الأقصى لأســعار الواردات ربع الســنوية، وأســعار التصدير التــي تقل عــن الحَدِّ الأدنى من أســعار التصدير ربع الســنوية. وينصــبُّ التركيز على أســعار التحويل غير الطبيعية بين الجلسات التي قد تؤدي إلى تحويل الدخل والضرائب

الخاضعة للضريبة من الولايات المتحدة. وقد يكون الانحراف المرصود في الأسعار ناتجًا عن تجنُّب/ تهرُّب ضريبة دخل، أو غسيل أموال أو تمويل إرهاب. وقد يرجع الانحراف السعري المرصود أيضًا إلى خطأ في قاعدة بيانات التجارة الأمريكية.

وجديرٌ بالذِّكر أنَّ التنقيب في البيانات سـوف يؤدي إلى تقييمٍ فعَّال للبيانات، وهو ما سيسـاعد بدوره في مكافحة الإرهاب. كما يمكن أن يُسـهِمَ تطبيق تقنية المعلومات وتقنيات التنقيب في البيانات في المعاملات المالية في تحسين المعلومات الاستخبارية.

أسئلة للمناقشة:

١- كيف يُمكِن استخدامُ التنقيب في البيانات لمكافحة الإرهاب؟ اذكر ما يُمكِن القيامُ به أيضًا غير ما تضمَّنته هذه الحالة العملية.

٢- هـل تعتقد أن التنقيب في البيانـات، رغم أنه ضروريٌّ لمحاربة الخلايا الإرهابية، يُهدِّد أيضًا حقوق الأفراد في الخصوصية؟

Sources: Zdanowic, J. S. (2004, May). Detecting money laundering and terrorist financing via data mining. Communications of the ACM, 47(5), 53; Bolton, R. J. (2002, January). Statistical fraud detection: A review. Statistical Science, 17(3), 235.

أسئلة للمراجعة على القسم ٤-٣:

١- ما هي مجالات التطبيق الرئيسة للتنقيب في البيانات؟

٢- حَدِّد على الأقل خمسة تطبيقات مُحدَّدة للتنقيب في البيانات، واذكر خمسة تصنيفات مشتركة لهذه التطبيقات.

٣- ما هو برأيك مجالُ التطبيق الأبرز للتنقيب في البيانات؟ ولماذا؟

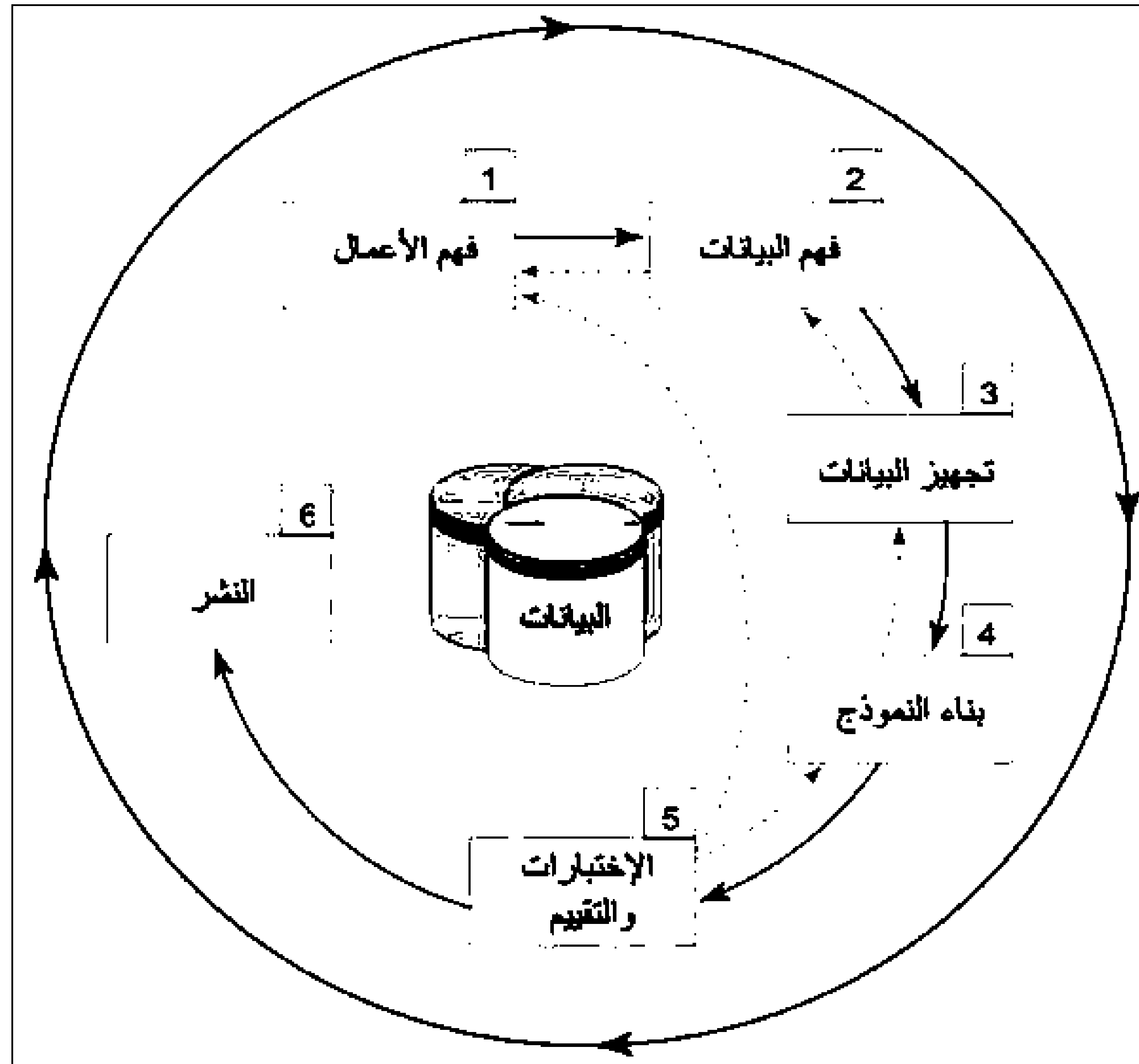
٤- هـل يُمكِنـك التفكيرُ في مجالات تطبيق أخـرى للتنقيب في البيانات لم تتم مناقشـتها في هذا القسم؟ وضِّح.

٤-٤ عملية التنقيب في البيانات:

عادةً ما يتمُّ اتباع طريقةٍ شائعة للقيام بمشاريع التنقيب في البيانات بشكلٍ منهجي. واستنادًا إلى أفضل الممارسات؛ قام الباحثون والممارسون في مجال التنقيب في البيانات باقتراح عِدة عمليات (سَـيْر عمل، أو مقاربات بسـيطة خطوة بخطوة)؛ لزيادة فرص النجاح في تنفيذ مشاريع التنقيب

في البيانات. وقد أدَّت هذه الجهود إلى العديد من العمليات القياسية، وفي هذا القسم سنذكر وَصْف بعضها (عددٌ قليلٌ من أكثر هذه العمليات شيوعًا).

وقد تمَّ اقتراحُ إحدى هذه العمليات الموحَّدة، التي يُمكِن اعتبارها الأكثر شيوعًا، وهي عمليةٌ قياسيةٌ عبر صناعة التنقيب في البيانات (CRISP-DM) في منتصف التسعينيات من قِبل الاتحاد الأوروبي للشركات للعمل كمنهجية غير مسجَّلة المِلكية للتنقيب في البيانات (CRISP-DM, 2013). ويوضِّح الشكل ٤-٣ هذه العملية المقترحة، وهي سلسلة من ست خطوات تبدأ بفَهْمٍ جيدٍ للأعمال والحاجة إلى مشروع التنقيب في البيانات (أيْ: مجال التطبيق)، وتنتهي بنشر الحل الذي يلبِّي المتطلبات المحَدَّدة لاحتياجات العمل. وعلى الرغم من أن هذه الخطوات متسلسلةٌ في الطبيعة، غير أنه عادةً ما يكون هناك قدرٌ كبيرٌ من التراجع. ونظرًا لأن التجربة والخبرة؛ هي التي تُحرِّك التنقيب في البيانات، وفقًا لحالة المشكلة والمعرفة/ خبرة المحلل؛ فإنه من الممكن أن تكون العملية برُمَّتها متكرِّرة للغاية (على سبيل المثال: يجب على المرء أن يتوقع أن يتنقل من خلال الخطوات عدة مرات) وتستغرق وقتاً طويلاً. ونظرًا؛ لأن الخطوات اللاحقة مبنيةٌ على نتائج الاختبارات السابقة؛ فإنه يجبُ على المرء أن يُولِي اهتمامًا إضافيًا للخطوات السابقة؛ من أجل عدم وَضْع الدراسة بالكامل على مسار غير صحيح منذ البداية.



شكل ٤-٣: عملية CRISP-DM ذات الست خطوات للتنقيب في البيانات

الخطوة ١- فهم الأعمال:

إن العنـصرَ الرئيس في أيِّ دراسـة للتنقيـب في البيانات؛ هو معرفة ماهية الدراسـة. وتبدأ الإجابة عن هذا السؤال بفَهْم شامل للحاجة الإدارية للمعرفة الجديدة ومواصفات صريحة لهدف العمل فيما يتعلق بالدراسـة التي سـيتمُّ إجراؤها. فنحن بحاجةٍ إلى أهداف مُحدَّدة مثل: «ما هي الخصائص المشـتركة للأفراد الذين فقدناهم لمنافسـتنا في الآونة الأخيرة؟» أو «ما هي الملامحُ النموذجية لأعواننا، وكم القيمة التي يوفرها كلٌّ منهم لنا؟». ثم يتمُّ تطويرُ خطة مشروع لإيجاد مثل هذه المعرفة التي تحَدِّد الأشـخاص المسـؤولين عن جمع البيانات، وتحليل البيانات، والإبلاغ عن النتائج. وفي هذه المرحلة المبكرة؛ يجب أيضًا إنشـاء ميزانية لدعم الدراسـة، على الأقل عند مستوى عالٍ مع أرقام تقريبية.

الخطوة ٢- فهم البيانات:

إنَّ دراسة التنقيب في البيانات مُخصَّصة للتصَدِّي لمهمة عمل مُحدَّدة تحديدًا جيدًا. وتتطلَّب مهامُّ العمل المختلفة مجموعاتٍ مختلفةً من البيانات. وبعد فهم الأعمال؛ فإن النشـاط الرئيس لعملية التنقيب في البيانات يتمثل في تحديد البيانات ذات الصِّلة من العديد من قواعد البيانات المتاحة. ويجبُ النظر في بعض النقاط الأساسية في مرحلة تحديد البيانات واختيارها. ويجبُ أولًا وقبـل كلِّ شيء، أن يكون المحللُ واضحاً وموجزًا حـول وَصْف مهمة التنقيب في البيانات؛ بحيث يُمكِن تحديدُ البيانات الأكثر صِلةً. فعلى سـبيل المثال: قد يسعى مشروعُ استخراج بيانات تجارة التجزئة إلى تحديد سلوكيات الإنفاق للمتسوِّقات الإناث اللاتي يقمن بشراء الملابس الموسمية على أساس التركيبة السـكانية، ومعاملات بطاقات الائتمان، والسِّمات الاجتماعية والاقتصادية. علاوةً عـلى ذلـك؛ يجبُ على المحلِّل بناءُ فهمٍ واعٍ لمصادر البيانات (مثل: أين يتمُّ تخزينُ البيانات ذات العلاقـة؟ وفي أيِّ شـكلٍ؟ وما هي عملية جَمْع البيانـات - الآلي مقابل اليدوي - ومَنْ هم جامعو البيانـات؟ وكـم مرةً يتمُّ تحديث البيانات؟) والمتغيرات (مثـل: ما هي المتغيرات الأكثر ملاءمةً؟ هل هناك أيُّ متغيِّرات مترادفة أو متشابهة؟ هل المتغيِّرات مستقلةً بعضها عن بعض؟ هل تقف كمصدر معلومات كامل دون تداخل أو تعارض معلومات؟).

ولفهم البيانات بشـكلٍ أفضل؛ فإنه غالبًا ما يسـتخدم المحللُ مجموعةً متنوعةً من التقنيات الإحصائية والرسـومية، مثل: ملخصات إحصائية بسـيطة لكلِّ متغير (فمثلًا: للمتغيرات الرقمية متوسطٌ يُشكِّل كلاً من الحَدِّ الأدنى أو الأقصى والوسيط والانحراف المعياري المقاييس المحسوبة؛ في

حين أنه بالنسـبة للمتغيرات الفئوية، فإنه يتمُّ حسابُ جداول الشكل والتردد)، وتحليل الارتباط، وقطاعات الانتشار، والمدرج الإحصائي، ومؤشرات الصندوق. إن التحديد الدقيق لمصادر البيانات واختيارها والمتغيرات الأكثر ملاءمةً يُمكِن أن تسهل على خوارزميات التنقيب في البيانات اكتشاف أنماط المعرفة المفيدة بسرعة.

ويُمكِـن أن تتنـوعَ مصادر البيانات؛ من أجـل انتقاء البيانات. ومن الناحيـة التقليدية؛ فإن مصـادر البيانات لتطبيقات الأعمال تشـمل كلاً مما يأتي: البيانـات الديموغرافية (مثل: الدخل، التعليـم، عدد المنازل، والعمر)، والبيانات الاجتماعية (مثل: الهوايات، وعضوية النادي، والترفيه)، وبيانات المعاملات (كسجل المبيعات، والائتمان إنفاق البطاقة، والشيكات المصدرة)، وما إلى ذلك. وفي الوقت الحاضر؛ فإن مصادر البيانات تستخدم أيضًا مستودعات البيانات الخارجية (المفتوحة أو التجارية)، ووسائط الإعلام الاجتماعية، والبيانات المُنشأة آليًّا.

كـما يُمكِن تصنيـف البيانات على أنهـا كميةٌ ونوعيـةٌ؛ إذ يتمُّ قياس الكمية باسـتخدام القيم الرقمية، أو البيانات الرقمية. وقد تكون منفصلةً (مثل: الأعداد الصحيحة) أو مستمرة (مثل الأرقام الحقيقيـة). وتحتوي البيانات النوعيـة، والمعروفة أيضًا بالبيانات الفئوية، على البيانات الاسـمية والترتيبية. فالبيانات الاسمية لها قيمٌ غير مرتبة مُحدَّدة (على سبيل المثال: بيانات النوع الاجتماعي، التي لها قيمتان: ذكورٌ وإناث). البيانات الترتيبية لديها قِيمٌ مُحدَّدة منتهية. على سبيل المثال: تُعَدُّ تقييـمات العملاء الائتمانية بيانات ترتيبية؛ لأن التقييمات يمكِن أن تكون ممتازة وعادلة كما يمكن أن تكون سيئةً. وقد وَرَدَ في الفصل ٢ تصنيفٌ بسيطٌ للبيانات (أيْ: طبيعة البيانات).

ويُمكِن تمثيلُ البيانات الكمية بسهولةٍ عن طريق نوعٍ من توزيع الاحتمالات؛ إذ يَصِف توزيعُ الاحتمالات كيف يتمُّ تشـتيت البيانات وتشـكيلها. فعلى سـبيل المثال: عادةً ما تكون البيانات الموزَّعة متماثلة ويُشـار إليها عادةً على أنها منحنى على شـكل جرس. كما يُمكِن تشفيرُ البيانات النوعية إلى أرقام، ثم وَصْفها من خلال توزيعات التردد. وبمجرد اختيار البيانات ذات الصِّلة وفقاً لهدف الأعمال التجارية للتنقيب في البيانات، يجب متابعة معالجة البيانات.

الخطوة ٣- إعداد البيانات:

إنَّ الغـرض من إعداد البيانات (والذي يُطلَق عليه عادةً اسـم المعالجة الأوَّلية للبيانات) هو أَخْذ البيانات المحَدَّدة في الخطوة السـابقة وإعدادها للتحليل بواسطة طرق التنقيب في البيانات. ومقارنـةً بالخطوات الأخرى في CRISP-DM؛ فإن المعالجة المُسـبَقة للبيانات تسـتهلك معظم الوقـت والجهد؛ ويعتقد الكثـيرون أنَّ هذه الخطوة تمثل ما يقرب مـن ٨٠٪ من إجمالي الوقت

الذي يقضيه في مشروع التنقيب في البيانات. والسبب في هذا الجهد الهائل الذي يتمُّ إنفاقه على هذه الخطوة، هو حقيقة أن بيانات العالم الحقيقي غير مكتملة بشكلٍ عام (أيْ: تفتقر إلى قِيم السِّمات، أو تفتقر إلى سِمات مُعيَّنة للاهتمام، أو تحتوي على بيانات مُجمَّعة فقط)، أو صاخبة (أيْ: تحتوي على أخطاء أو أخطاء خارجية)، وغير متناسقة (أيْ: تحتوي على تناقضات في الرموز أو الأسماء). وقد ورد شرحُ طبيعة البيانات والمسائل المتعلِّقة بعمليات المعالجة المُسبَقة لبيانات التحليلات بالتفصيل في الفصل الثاني.

الخطوة ٤- بناء النماذج:

وفي هذه الخطوة يتمُّ اختيارُ تقنيات النمذجة المختلفة وتطبيقها على مجموعة بيانات تمَّ إعدادُها بالفعل لتلبية احتياجات العمل المحَدَّدة. وتشتمل خطوة بناء النموذج أيضًا على التقييم والتحليل المقارن لمختلف النماذج المبنية. ونظرًا لعدم وجود طريقة أو خوارزمية أفضل معروف عالميًّا لمهمة التنقيب في البيانات؛ فإنه يجبُ على المستخدِم استخدامُ مجموعة متنوعة من أنواع النماذج القابلة للتطبيق جنباً إلى جنب مع إستراتيجية تقييم وتجربة مُحدَّدة تحديدًا جيدًا لتحديد الطريقة «الأفضل» لغرض معين. وحتى بالنسبة لطريقة واحدة أو خوارزمية واحدة؛ فإنه يجبُ تحديدُ عددٍ من المعلمات للحصول على أفضل النتائج. وقد تحتوي بعضُ الأساليب على متطلبات مُحدَّدة في الطريقة التي يتمُّ بها تنسيق البيانات؛ وبالتالي؛ فإن العودة إلى خطوة إعداد البيانات تكون ضروريةً في كثير من الأحيان. وتقدِّم الحالة العملية ٤-٤ دراسةً بحثيةً يتمُّ فيها تطوير عددٍ من أنواع النماذج ومقارنتها بعضها ببعض.

وتبعًا لاحتياجات العمل؛ فإنه يُمكِن أن تكون مهمة التنقيب في البيانات للتنبؤ (إما التصنيف أو الانحدار) أو اقتران أو نوع تجميع. كما يُمكِن لكلٍّ من هذه المهام التنقيب في البيانات من خلال استخدام مجموعةٍ متنوعةٍ من الخوارزميات وأساليب التنقيب في البيانات. وقد تناولنا شرحَ بعض طرق التنقيب في البيانات في جزءٍ سابق من هذا الفصل، وسوف نتناول لاحقًا في هذا الفصل وَصْف بعض أكثر الخوارزميات استخدامًا، بما في ذلك شجرات القرار للتصنيف، وk-means لتجمع الحلول، وخوارزمية Apriori لاستخراج قواعد الارتباط.

الخطوة ٥- الاختبار والتقييم:

في هذه الخطوة؛ يتمُّ تحديدُ وتقييمُ النماذج المتقدِّمة من حيث دقتها وجدارتها؛ إذ تقوم هذه الخطوة بتقييم الدرجة التي يُلبِّي بها النموذج (أو النماذج) المحَدَّدة أهداف العمل، وإذا كان الأمرُ

كذلك؛ فإلى أيٍّ مدى (أيْ: هل هناك حاجةٌ إلى تطوير المزيد من النماذج وتقييمها؟). وهناك خيارٌ آخر، وهو اختبارُ النموذج (أو النماذج) المتقدِّم في سيناريوهات العالم الحقيقي في حالةِ ما إذا كانت قيود الوقت والميزانية تسـمح. وعلى الرغم من أنه مـن المتوقع أن ترتبط نتائج النماذج المتقدِّمة بأهـداف العمل الأصلية؛ فإنَّ النتائج الأخرى التي لا ترتبـط بالضرورة بأهداف العمل الأصلية، قد تكشف النقاب عن معلومات إضافية أو تلميحات للاتجاهات المستقبلية التي يتمُّ اكتشافها.

وتُعَـدُّ خطـوة الاختبار والتقييم مهمـةً حرجةً وصعبةً؛ إذ يتمُّ إضافة أيٍّ قيمة بواسـطة مهمة التنقيب في البيانات؛ حتى يتمَّ التعرُّف على قيمة الأعمال التي تمَّ الحصولُ عليها من أنماط المعرفة المُكتشـفة والتعرُّف عليها. فتحديد القيمة التجارية من أنماط المعرفة المُكتشـفة يشـبه إلى حَدٍّ ما لعب الألغاز. وأنماط المعرفة المستخلصة؛ هي أجزاءٌ من اللغز والتي يجبُ أن يتمَّ تجميعها في سياق غرض العمل المحَدَّد. ويعتمد نجاح عملية تحديد الهوية على التفاعل بين مُحللي البيانات ومُحلِّلي الأعمال وصانعي القرار (مثل مديري الأعمال). ونظرًا لأن مُحلِّلي البيانات قد لا يكون لديهم الفهم الكامـل لأهداف التنقيـب في البيانات وما يعنيه لرجال الأعـمال، ومُحلِّلي الأعمال، كما أن صانعي القرار قد لا يكون لديهم المعرفة التقنية لتفسير نتائج الحلول الرياضية المعقدة؛ ولذلك فإن التفاعل بينهم أمرٌ ضروري. ولتفسـير أنماط المعرفة بشـكلٍ صحيح؛ غالبًا ما يكون من الضروري اسـتخدام مجموعـةٍ متنوعةٍ من تقنيات الجدولة والتصور (على سـبيل المثال: الجـداول المحورية، الجدولة المتقاطعة للنتائج، المخططات الدائرية، المدرج التكراري، مخططات الأرض، نقاط التشتت).

الخطوة ٦- النَّشر:

إن تطويرَ وتقييمَ النماذج ليس نهاية مشروع التنقيب في البيانات. وحتى إن كان الغرض من هذا النموذج هو الحصول على استكشافٍ بسيط للبيانات؛ فإن المعرفة المُكتَسَبة من مثل هذا الاستكشاف ستحتاج إلى تنظيم وتقديمها بطريقةٍ تجعل المستخدِم النهائي قادرًا على فهمها والاستفادة منها. وبناءً على المتطلبات؛ فإن مرحلة النشر قد تكون بسـيطةً مثل إنشـاء تقرير أو معقدة، مثل تنفيذ عملية التنقيـب في البيانات القابلة للتكرار عبر المؤسسـة. وفي العديدِ مـن الحالات؛ يكون العميلُ، وليس مُحلِّل البيانات، هو الذي يُنفِّذ خطوات النشر. ومع ذلك فحتى إذا لم يقم المُحلِّل بتنفيذ جهد النشر؛ فمن المهمِّ أن يفهم العميلُ في البداية ما هي الإجراءات التي يتعَيَّن القيام بها لاستخدام النماذج التي تم إنشـاؤها فعليًّا. وقد تتضمَّن خطوة النشر أيضًا أنشـطة الصيانة للنماذج التي تمَّ نشرها. نظرًا لأن كلَّ شيءٍ يتعلق بالأعمال يتغير باسـتمرار؛ فإن البيانات التي تعكسُ أنشـطة الأعمال تتغير أيضًا. وبمـرور الوقت؛ قد تصبح النماذج (والأنماط المضمَّنة بها) المبنية على البيانات القديمة عتيقةً أو غير

ذات صِلـة أو مضللـة. ولذلك فمن المهـمِّ مراقبة النماذج وصيانتهـا إذا أصبحت نتائج التنقيب في البيانات جزءًا من الأعمال اليومية وبيئتها. ويساعد الإعدادُ الدقيق لإستراتيجية الصيانة على تجنُّب فـترات طويلة غير ضرورية من الاسـتخدام غير الصحيح لنتائج التنقيـب في البيانات. ولمراقبة نشر نتيجـة (نتائج) التنقيب في البيانات، يحتاج المشروع إلى خطة مُفصَّلة حول عملية المراقبة، والتي قد لا تكون مهمةً بسيطةً بالنسبة لنماذج التنقيب في البيانات المعقدة.

حالة عملية ٤-٤

يساعد التنقيب في البيانات في أبحاث السرطان

وفقاً لجمعية السرطان الأمريكية؛ فإن نصف الرجال وثلث النساء في الولايات المتحـدة الأمريكيـة سـيصابون بالسرطان خـلال حياتهم؛ ومن المتوقـع أن يتمَّ تشخيص ١,٥ مليون حالة سرطان جديدة في عام ٢٠١٣م. ويُعَدُّ السرطان هو ثاني أكثر أسـباب الوفاة شـيوعًا في الولايات المتحدة والعالم، ولا يتجاوزه سوى أمراض القلـب والأوعيـة الدموية فقط. ومـن المتوقع في هذا العـام أن يموت أكثر من ٥٠٠,٠٠٠ أمريكي بسـبب السرطان - وهذا يعني أكثر من ١٣٠٠ شخص في اليوم - وهو ما يمثل حالة وفاة واحدة من كلِّ أربع حالات.

والسرطان هو مجموعةٌ من الأمراض يتمُّ تمثيلها بشـكلٍ عام عن طريق النمو غـير المنضبط وانتشـار الخلايا غير الطبيعية. وإذا لم يتـم التحكُّم في النمو و/ أو الانتشار؛ فقد يؤدي ذلك إلى الوفاة. على الرغم من عدم معرفة الأسباب الدقيقة، ومن المُعتقد أن السرطان ناتجٌ عن عوامل خارجية (مثل: التبغ، والكائنات المعدية، والكيميائية، والإشعاعية) والعوامل الداخلية (مثل: الطفرات الوراثية، والهرمونات والظـروف المناعية والطفرات التي تحدث بسـبب التمثيل الغذائي). وقد تعمل هذه العوامل السَّـببية معًا أو في تسلسلٍ لبَدْء أو تعزيز التسرطن. ويتم التعامل مع السرطان بالجراحة، والإشعاع، والعلاج الكيميائي، والعلاج بالهرمونات، والعلاج البيولوجي، والعلاج الموجَّه. وتختلف إحصاءات البقاء على قيد الحياة بشكلٍ كبيرٍ حسب نوع السرطان ومرحلة التشخيص.

إنَّ معـدَّل البقاء على قيد الحياة لمدة ٥ سـنوات لجميـع السرطانات آخذٌ في التحسُّـن، كـما أن معدل وفيات السرطـان قد انخفـض إلى ٢٠٪ في عام ٢٠١٣م،

وهو مـا أدَّى إلى تجنُّـب حـوالي ١,٢ مليـون حالة وفـاة بسـبب السرطان منذ عام ١٩٩١م. ويعني هذا أن أكثر من ٤٠٠ شـخص يتمُّ إنقاذهم يوميًّا! ويعكس التحسُّـن الـذي طرأ على عددِ الأشـخاص الذين يتـمُّ إنقاذهم مـدى التقدُّم في تشـخيص بعض أنواع السرطان في مرحلة مبكرة والتحسـينات في العلاج. غير أننا ما زِلنا بحاجة إلى مزيدٍ من التحسينات لمنع وعلاج السرطان.

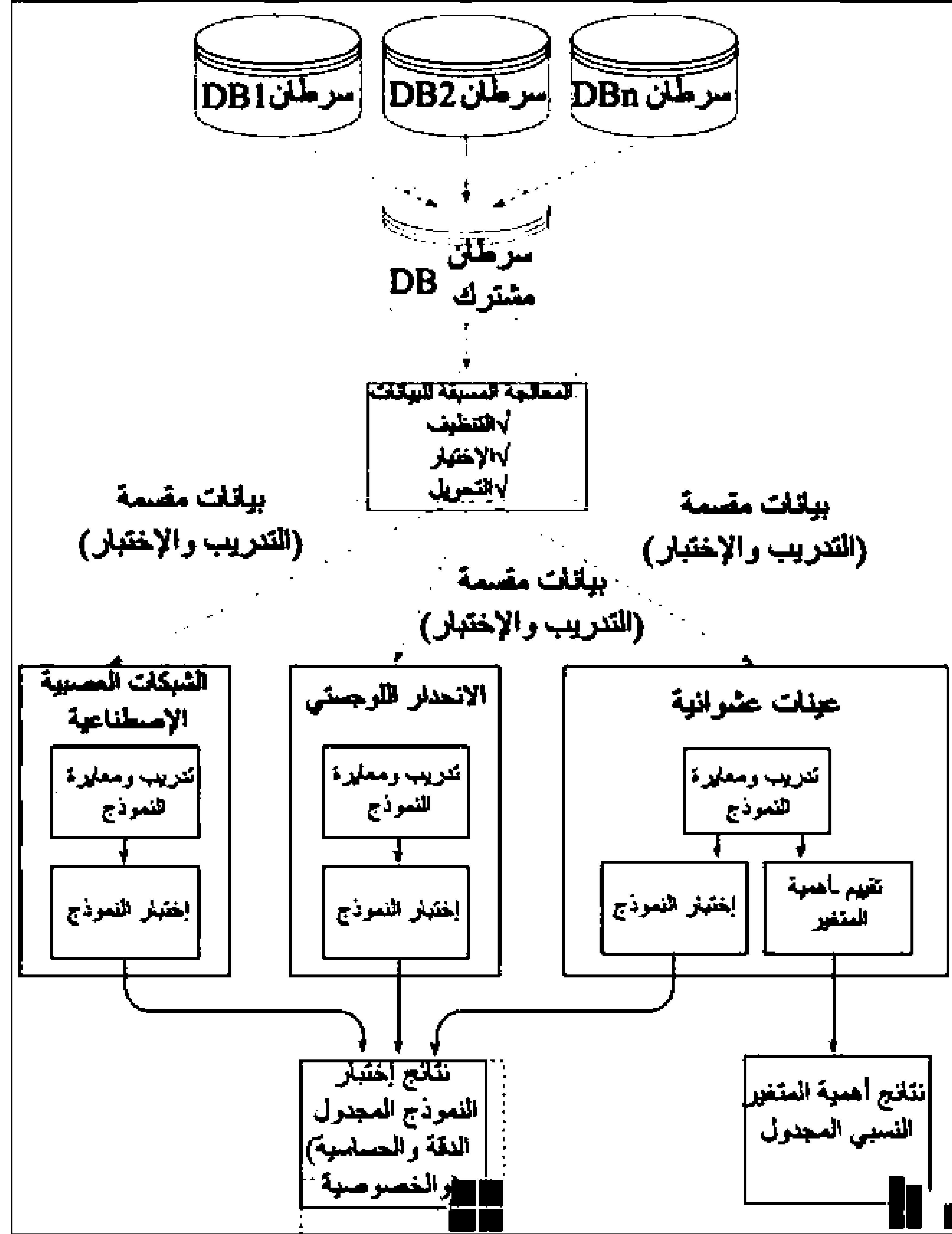
وعـلى الرَّغم من أن أبحاث السرطان كانـت ذات طابعٍ سريريٍّ وبيولوجي في العادة؛ فإن الدراسـات التحليلية المسـتندة إلى البيانات أصبحت مُكمِّلًا شائعًا في السنوات الأخيرة. وقد تمَّ تحديدُ اتجاهات البحوث الجديدة في المجالات الطبية؛ إذ تـمَّ تطبيق الأبحـاث التي تعتمدُ على البيانات والتحليـلات بنجاح، وذلك من أجل تعزيز الدراسـات السريريـة والبيولوجية. وقد تمكَّـن الباحثون من تحديد أنماط جديدة باستخدام أنواع مختلفة من البيانات، بما في ذلك البيانات التجريبية الجزيئيـة، والسريريـة، والمرتكـزة على الأدبيـات (الكتابات التـي كُتبت في هذا الموضـوع)، ومعلومات التجارب السَّريرية، إلى جانب أدوات وتقنيات التنقيب في البيانات المناسبة، وقد تمكَّن الباحثون من تحديد أنماط جديدة؛ مما يمهِّد الطريق نحو مجتمعٍ خالٍ من السرطان.

وفي إحدى الدراسـات استخدم Delen (2009) ثلاث تقنيات شائعة للتنقيب في البيانات؛ وهي: (أشجار القرار، والشبكات العصبية الاصطناعية، وآلة المتجهات الداعمة) بالتزامن مع الانحدار اللوجستي (المنطقي) لتطوير نماذج التنبؤ للقدرة على البقاء مع وجود سرطان البروسـتاتا. وقـد تضمَّنَت مجموعة البيانات حوالي ١٢٠,٠٠٠ سجل و٧٧ متغيرًا. كما تمَّ استخدام منهجية التحقق المتقاطع س - جزء (K-fold) في بنـاء النماذج والتقييـم والمقارنة. وقد أوضحـت النتائجُ أن نماذج المتجهـات الداعمة؛ هـي الأكثر دقة (مع دقة ضبط تبلـغ ٩٢,٨٥٪) لهذا المجال، تليها الشبكات العصبية الاصطناعية وأشجار القرار. وعلاوةً على ذلك؛ وباستخدام طريقة تقييمٍ تعتمد على الحساسـية - التحليلية، كشفت الدراسة أيضًا عن أنماط جديدة تتعلق بالعوامل التي تُنذِر بوجود سرطان البروستاتا.

وفي دراسةٍ ذات صِلة استخدم فيها كلٌّ من Delen، وWalker، و Kadam (2005)، اثنـين من خوارزميات اسـتخراج البيانات، وهي: (الشـبكات العصبية الاصطناعية،

وأشجار القرار) والانحدار اللوجستي (المنطقي) لتطوير نماذج التنبؤ لبقاء سرطان الثدي باستخدام مجموعةٍ كبيرة من البيانات (أكثر من ٢٠٠,٠٠٠ حالة). وباستخدام منهجية التحقق المتقاطع المكوَّنة من ١٠ أجزاء لقياس التقدير غير المتحيِّز لنماذج التنبؤ لأغراض مقارنة الأداء - أشارت النتائجُ إلى أن شجرة القرارات (خوارزمية C5) كانت أفضل طريقة تنبؤ، مع دقة ٩٣,٦٪ على العينة المخالفة (والتي كانت أفضل طريقة تنبؤ ذُكِرت في الأدبيات)، تليها الشبكات العصبية الصناعية، مع دقة ٩١,٢٪، والانحدار اللوجستي (منطقي)، مع دقة ٨٩,٢٪. وقد كشف المزيدُ من التحليل لنماذج التنبؤ عن أهمية تنبؤات العوامل، والتي يمكن استخدامها بعد ذلك كأساس لمزيدٍ من الدراسات البحثية السريرية والبيولوجية.

وفي الدراسة الأخيرة؛ قام كلٌّ من Zolbanin، وDelen، وZadeh (2015) بدراسة تأثير الاعتلال المشترك في البقاء على قَيْد الحياة للسرطان. وعلى الرغم من أن الأبحاث السابقة أظهرت أن التوصيات التشخيصية والعلاجية يمكِن أن تتغير بناءً على شِدة الأمراض المصاحبة، غير أن الأمراض المزمنة لا تزال قيد البحث بمعزل بعضها عن بعض في معظم الحالات. ولتوضيحِ أهمية الأمراض المزمنة المتزامنة في أثناء العلاج؛ فقد استخدمت دراستهم المراقبة، وعلم الأوبئة، والنتائج النهائية (SEER) لإنشاء مجموعتين من البيانات المَرضية: إحداهما لسرطانات الثدي والأعضاء التناسلية للإناث، والأخرى لسرطانات البروستاتا والسرطانات البولية. ثم يتمُّ تطبيق العديد من تقنيات تعلُّم الآلة الشائعة على مجموعات البيانات الناتجة لبناء نماذج تنبؤية (انظر: شكل ٤-٤). وقد أظهرت مقارنة النتائج أن الحصول على مزيدٍ من المعلومات حول الحالات المرضية للمرضى يمكن أن يحسِّن من القدرة التنموية لدى النماذج، والتي بدورها يمكِن أن تساعد الممارسين على اتخاذ قرارات أفضل في التشخيص والعلاج. ولذلك؛ فقد اقترحت الدراسة أن التحديد السليم وتسجيل واستخدام حالة الاعتلال المشترك للمرضى يمكِن أن يقلل تكاليف العلاج ويخفف من التحَدِّيات الاقتصادية المتعلقة بالرعاية الصحية.



وتُظهِر هذه الأمثلة (ضِمن العديدِ من الدراسات الأخرى في الأدبيات الطبية) أنه يُمكِن اســتخدام تقنيات دقيقة للبيانات المتقدِّمة لتطوير نماذج تمتلك درجةً عاليةً من القدرة التنبؤية والتفســيرية. وعلى الرغم من أن أساليب التنقيب في البيانات قادرةٌ على اســتخراج الأنماط والعلاقات المخبَّأة في أعــماق قواعد البيانات الطبية الكبــيرة والمعقَّدة، دون التعاون والتغذية المرتدة من الخبراء الطبيين؛ فإن نتائجها ليســت مفيدةً بشكلٍ كبير؛ إذ يجبُ تقييم الأنماط الموجودة عبر أساليب التنقيب في البيانات من قِبل المتخصِّصين في المجال الطبي الذين لديهم ســنوات من الخبرة في مجال المشكلة؛ لتحديد ما إذا كانت منطقيةً وقابلةً للتطبيق ومبتكرةً بما يكفي

لتبرير اتجاهات بحثية جديدة. وباختصار؛ فإنه لا يُقصَد من التنقيب في البيانات أن تحــلَّ محلَّ المهنيين والباحثين المتوســطين؛ بل تجمع جهودهــم التي لا تُقدَّر بثمن لتوفير اتجاهات بحثية جديدة مدفوعة بالبيانات ولإنقاذ المزيد من الأرواح البشرية في نهاية المطاف.

أسئلة للمناقشة:

١- كيف يُمكِن اســتخدامُ التنقيب في البيانات لعلاج الأمراض في نهاية المطاف مثل السرطان؟

٢- مــا هي في اعتقادك الوعود والتحَدِّيــات الكبرى التي يواجهها العاملون في مجال التنقيب في البيانات للمساهمة في الجهود البحثية الطبية والبيولوجية؟

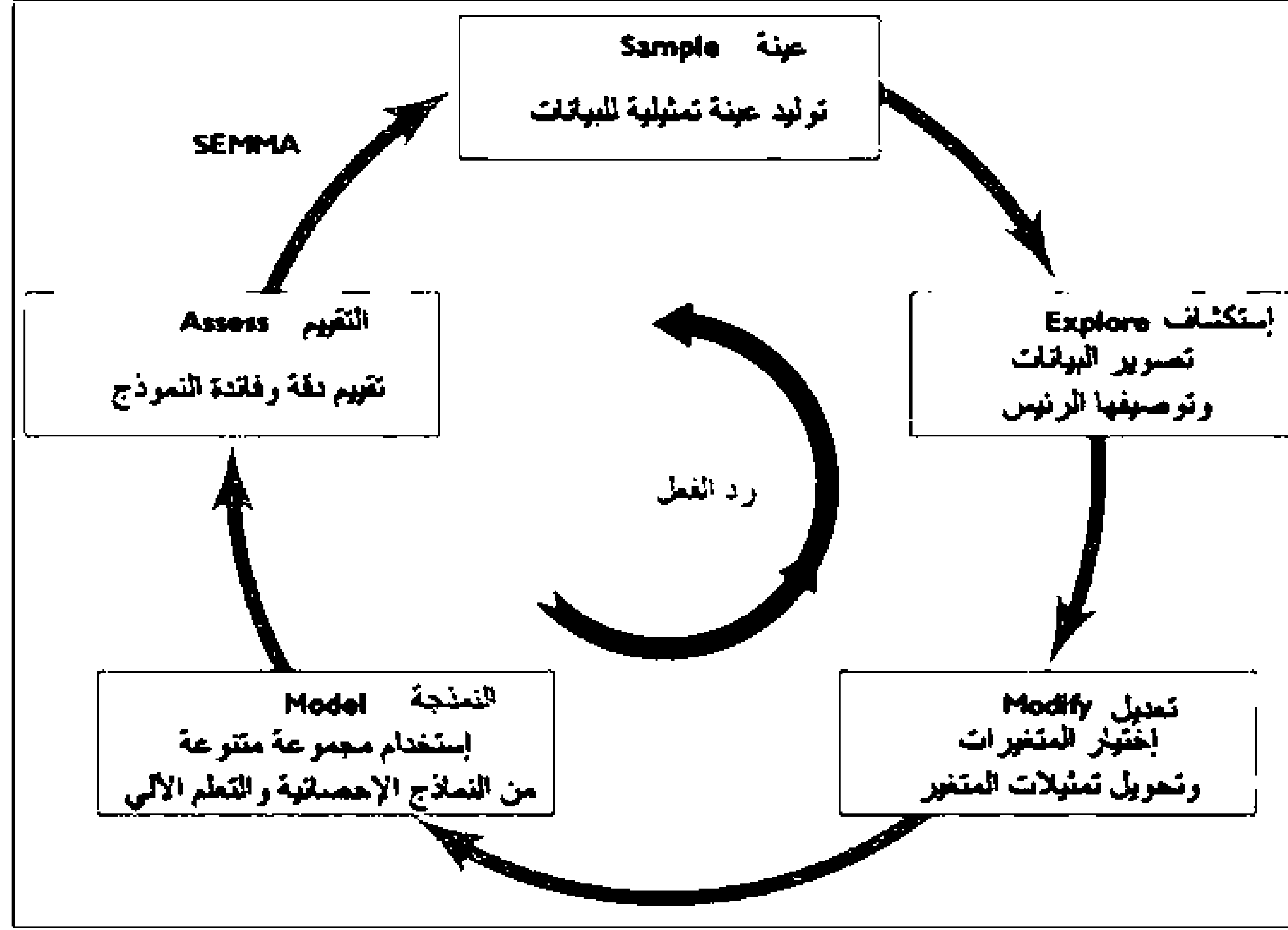
Sources: Zolbanin, H. M., Delen, D., & Zadeh, A. H. (2015). Predicting overall survivability in comorbidity of cancers: A data mining approach. Decision Support Systems, 74, 150161-; Delen, D. (2009). Analysis of cancer data: A data mining approach. Expert Systems, 26(1), 100112-; Thongkam, J., Xu, G., Zhang, Y., & Huang, F. (2009). Toward breast cancer survivability prediction models through improving training space. Expert Systems with Applications, 36(10), 1220012209-; Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. Artificial Intelligence in Medicine, 34(2), 113127-.

عمليات ومنهجيات أخرى مُوحَّدة للتنقيب في البيانات:

ولكي يتمَّ تطبيقها بنجاح؛ فإنه يجبُ النظر إلى دراسة التنقيب في البيانات على أنها عملية تتبع منهجيــة واحدة بدلًا من مجموعة من أدوات وتقنيات البرامج الآلية. فإضافةً إلى CRISP-DM؛ هناك منهجيةٌ أخرى معروفة تمَّ تطويرها من قِبل معهد SAS، وتُســمَّى SEMMA (2009)؛ إذ يشيرُ اختصار SEMMA إلى «العينة، والاستكشاف، والتعديل، والطراز، والتقييم».

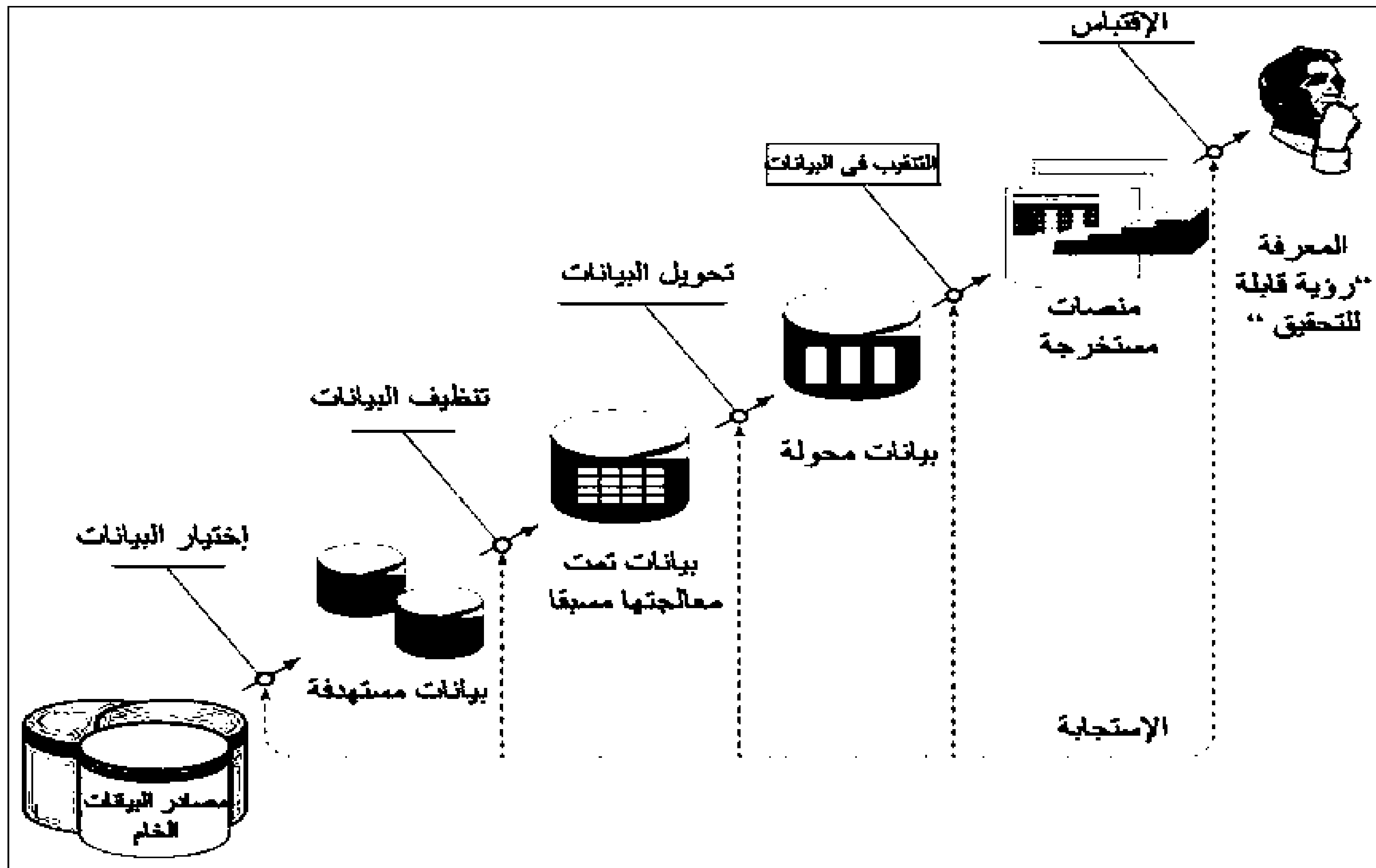
وبــدءًا من عيِّنة مــن البيانات تمَّ تمثيلها إحصائيًّا؛ فإنَّ SEMMA يجعلُ من الســهل تطبيق الأســاليب الاستكشــافية الإحصائية والتصوير، وتحديد وتحويل أهمِّ المتغــيِّرات التنبؤية، وتهيئة المتغيِّرات للتنبؤ بالنتائج، وتأكيد دقة النموذج. وهناك تمثيلٌ مُصوَّر لـ SEMMA في الشكل ٤-٥. ومــن خلال تقييم نتائج كلِّ مرحلة من مراحل عمليــة SEMMA؛ يمكن لمطوِّر النموذج تحديد كيفيــة صياغة أســئلة جديدة تثيرها النتائج الســابقة. وبالتالي، العودة إلى مرحلة الاستكشــاف

من أجل تحســين البيانات بشــكلٍ أفضل؛ وذلك كما هو الحال مع CRISP-DM، إذ يتمُّ تشغيلُ SEMMA بدورةٍ تجريبية متكررة للغاية. ويتمثل الفرق الرئيس بين CRISP-DM، وSEMMA في أن CRISP-DM يتخذ منهجًا أكثر شــموليةً - بما في ذلك فهم الأعــمال والبيانات ذات الصِّلة - بالقيــاس بمشروعات التنقيب في البيانات، في حين يفترض SEMMA ضمنيًّا أن أهداف وغايات مشروع التنقيب في البيانات تكون معًا جنبًا إلى جنب مع تحديد مصادر البيانات المناسبة وفهمها.



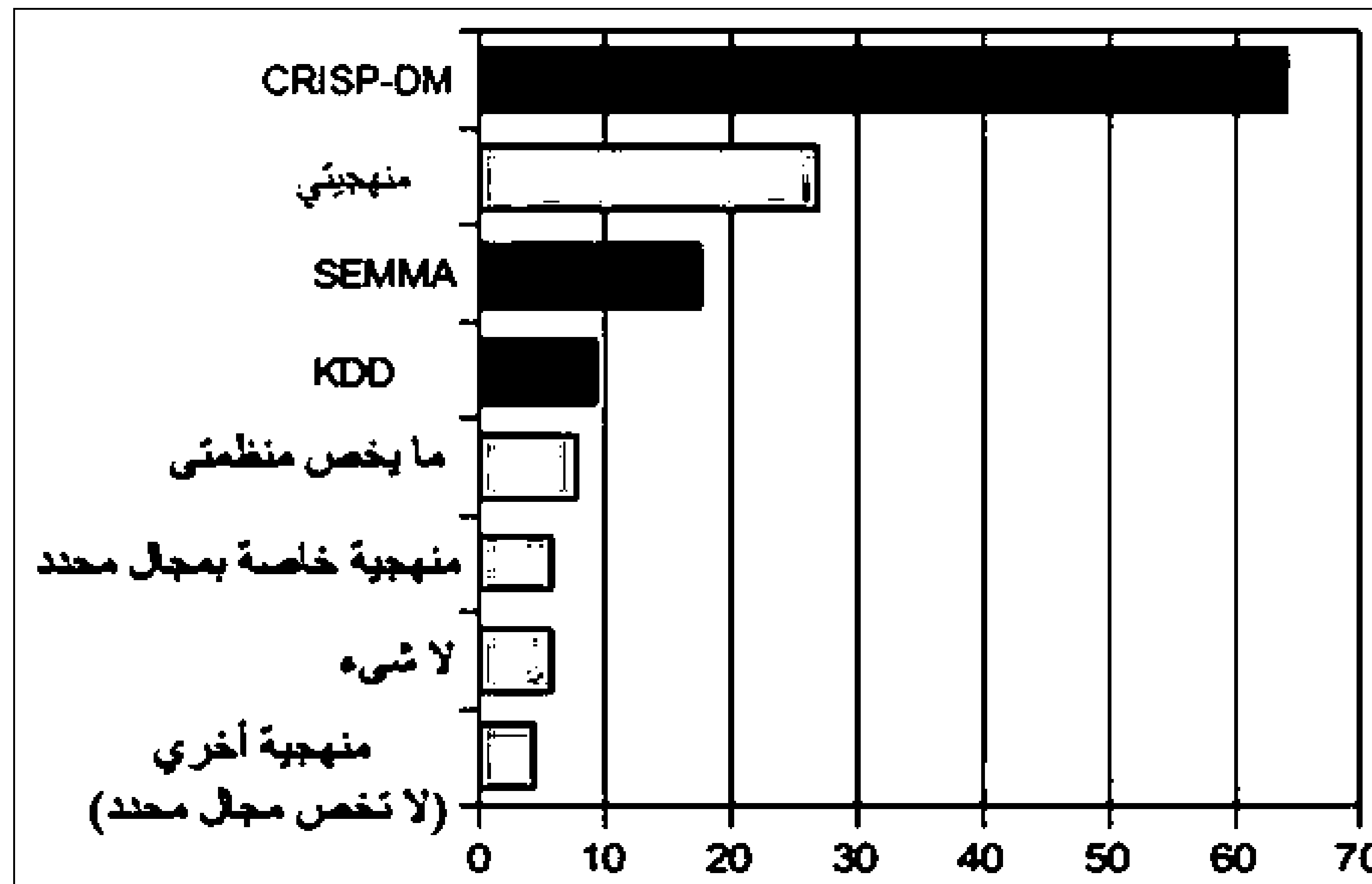
شكل ٤-٥: عملية SEMMA للتنقيب في البيانات

وعادةً ما يَســتخدِم بعض الممارســين مصطلح اكتشــاف المعرفة في قواعد البيانات (KDD) كمــرادفٍ للتنقيب في البيانات. وقد عرَّف Fayyad وآخرون (١٩٩٦) اكتشــاف المعرفة في قواعد البيانات بأنها عمليةٌ لاستخدام أساليب التنقيب في البيانات؛ للعثور على معلومات وأنماط مفيدة في البيانات، على عكسِ تعدين البيانات، والتي تنطوي على استخدام الخوارزميات لتحديد الأنماط في البيانات المشــتقة؛ من خلال عملية KDD (انظر: الشــكل ٤-٦). حيث إنَّ KDD هي عمليةٌ شــاملةٌ تشــمل التنقيب في البيانات، ويتكون الإدخال إلى عملية KDD من بيانات تنظيمية، كما يتيح مســتودع بيانات للمؤسسة تنفيذ KDD بكفاءة؛ لأنه يوفر مصدرًا واحدًا للبيانات التي يتمُّ اســتخراجها. وقد لخَّص Dunham (2003) عملية KDD على أنها تتكوَّن من الخطوات التالية: اختيار البيانات، معالجة البيانات، تحويل البيانات، التنقيب في البيانات، التفسير/ التقييم.



شكل ٤-٦: عملية KDD (اكتشاف المعرفة في قواعد البيانات)

ويوضِّح الشــكل ٤-٧ نتائج الاستطلاع للســؤال، «ما هي المنهجية الرئيسة التي تستخدمها في التنقيب في البيانات؟» (تم الاستطلاع عن طريق kdnuggets.com في أغسطس ٢٠٠٧).



Source: Used With Permission From Kdnuggets.com

شكل ٤-٧: رتب منهجيات/ عمليات التنقيب في البيانات

أسئلة للمراجعة على قسم ٤-٤:

١- ما هي عمليات التنقيب في البيانات الرئيسة؟

٢- لماذا تعتقدُ أنَّ المراحل الأولى (فهم العمل وفهم البيانات) هي الأطول في مشاريع التنقيب في البيانات؟

٣- اذكر مراحل عملية CRISP-DM، وعرِّفها بإيجاز.

٤- ما هي الخطواتُ الرئيســة لمعالجة البيانات؟ قم بوَصْف كلِّ خطوة بإيجاز، مع تقديم أمثلةٍ ذات صِلة.

٥- ما هو الفرق بين CRISP-DM، وSEMMA؟

٤-٥ طرق التنقيب في البيانات:

تتوفَّر العديدُ من الطرق المتنوعة لأداء دراسات التنقيب في البيانات، والتي تشمل: التصنيف، والانحدار، والتجميع، والترابط. وتَستخدِم معظمُ أدوات برامج التنقيب في البيانات أكثر من تقنيةٍ (خوارزمية) لكلٍّ من هذه الطرق. وفي هذا القسم نقدِّم وصفًا لأساليب التنقيب في البيانات الأكثر شيوعًا مع شرح التقنيات التي تقوم بتمثيلها.

التصنيف:

ربمــا تكون طريقة التصنيف للتنقيب في البيانات؛ هي الطريقة الأكثر اســتخدامًا لمشــكلات العالم الحقيقي. وبصفتها عضوًا مشــهورًا في عائلة تقنيات العالم الآلي؛ فإنَّ التصنيف يتعلم أنماطًا من البيانات الســابقة (مجموعة من المعلومات - السِّمات والمتغيِّرات والميزات - على العناصر أو الكائنات أو الأحداث المسماة سابقًا) لوضع حالات جديدة (مع بطاقات تعريف غير معروفة) في مجموعاتهم أو فصولهم. فعلى ســبيل المثال: يُمكِن استخدام التصنيف للتنبؤ بما إذا كان الطقس في يومٍ معين سيكون «مشمسًا» أو «ممطرًا» أو «غائمًا»، وتتضمَّن مهامُّ التصنيف الشائعة الموافقة على الائتمان (أيْ: المخاطر الائتمانية الجيدة أو الســيئة)، وموقع المتجر (أيْ: هل الموقع جيد، أو معتدل، أو ســيئ؟)، والتســويق المُستَهدَف (على ســبيل المثال: العميل المُحتَمَل، أو الذي لا أمل منه)، وكشف الاحتيال (أيْ: نعم/ لا)، والاتصالات السِّلكية واللاسلكية (بمعنى أنه من المحتمل أن يتحوَّل إلى شركة هاتف أخرى، نعم/ لا). فإذا كان المتوقع هو تســمية فئة (مثل: «مشــمس» أو «ممطر» أو «غائم»)؛ فإن مشكلة التنبؤ تُسمَّى تصنيفاً؛ في حين إذا كانت قيمة رقمية (على سبيل المثال: درجة الحرارة، مثل: ٦٨ درجة F)؛ فإن مشكلة التنبؤ تُسمَّى انحداراً.

وعـلى الرَّغم من أنه يُمكِن اسـتخدامُ التجميع (وهـي طريقةٌ أخرى شـائعةٌ للتنقيب في البيانات)؛ لتحديد المجموعات (أو عضوية الصَّف) للأشـياء؛ فهناك فـرق كبيرٌ بين الاثنين؛ إذ يتعرَّف التصنيف على الوظيفة بين خصائص الأشياء (أيْ: المتغيِّرات المستقلة) وعضويتها (أيْ: متغيِّر المخرجات) من خلال عملية التعلُّم تحت الإشراف؛ إذ يتمُّ تقديمُ كلا النوعين (المدخلات والمخرجات) من المتغيرات إلى خوارزمية؛ أمَّا في التجميـع؛ فإن تعلُّم عضوية الكائنات يتمُّ من خـلال عملية تعلُّم غير خاضعة للإشراف؛ إذ يتمُّ عَرْض مُتغـيِّرات الإدخال فقـط على الخوارزمية. وعلى عكـس التصنيف؛ فإن التجميع ليس لـه آليةٌ إشرافيةٌ (أو مراقبـة) تفرض عملية التعلُّم، وبدلًا من ذلك؛ فإن الخوارزميات الخاصة بالتجميع تسـتخدم واحدًا أو أكثر من الاستدلالات (مثل: قياس المسافة مُتعدِّد الأبعاد) لاكتشاف التجمعات الطبيعية للأجسام.

وجديـرٌ بالذِّكـر أنَّ منهجية الخطوتين الأكثر شـيوعًا للتنبؤ بنوع التصنيـف تتضمَّن تطويراً/ تدريباً نموذجياً، واختباراً / نشراً نموذجياً. وفي مرحلة تطوير النموذج؛ يتمُّ استخدامُ مجموعةٍ من بيانات الإدخال، بما في ذلك ملصقات الفئات الفعلية. وبعد أن يتمَّ تدريبُ النموذج، يتم اختبار النموذج مقابل عينة الحجز لتقييم الدقة، وفي النهاية يتمُّ نشره للاسـتخدام الفعلي؛ إذ يتمُّ التنبؤ بفئـات مثيلات البيانات الجديدة (إذ يكون تصنيف الفئـة غير معروف). وعند تقييم النموذج؛ فإنَّ هناك عدة عوامل لا بُدَّ من أخذها في الاعتبار، وتتضمَّن الآتي:

- **الدِّقة التنبؤية:** هي قدرة النموذج على التنبؤ بشكلٍ صحيحٍ بتصنيف الفئة للبيانات الجديدة أو غير المرئية مسبقاً. ودقة التنبؤ هي عامل التقييم الأكثر استخدامًا لنماذج التصنيف. ولحساب هذا المقياس؛ فإنه يتمُّ مطابقة تصنيفات الطبقة الفعلية لمجموعة بيانات اختبار مع تصنيفات الفئات التي يتنبأ بها النموذج. وبعد ذلك يُمكِن احتسـابُ الدقة كمعدل دقة، وهي النسـبة المئويـة لعيِّنات مجموعة بيانات الاختبار التي تمَّ تصنيفها بشـكلٍ صحيح بواسـطة النموذج (وسوف نتناول هذا الموضوع بشيءٍ من التفصيل لاحقًا في الفصل).
- **السرعة:** هي التكاليف الحسـابية التي ينطوي عليها توليدُ واستخدام النموذج، وكلما كان ذلك أسرع؛ فإنه يُعَدُّ أفضل.
- **المتانة:** هي قدرة النموذج على إجراء تنبؤات دقيقة إلى حَدًّ معقول عندما يتمُّ إعطاؤه بيانات صاخبة (غير دقيقة)، أو بيانات ذات قيم مفقودة أو خطأ.
- **قابلية التوسُّع:** هي القدرة على بناء نموذج التنبؤ بكفاءة تُعطِى كميةً كبيرةً من البيانات.
- **التفســير:** هو مسـتوى الفهم والبصيرة اللـذان يوفرهما النموذج (مثل: كيـف أو ما يختتمه النموذج في بعض التنبؤات).

تقدير الدِّقة الحقيقية لنماذج التصنيف:

يُعَدُّ المصدرُ الأســاسي لتقدير الدِّقة في مشكلات التصنيف؛ هو مصفوفة الدُّقة (وتُسمَّى أيضًا مصفوفة التصنيف أو جدول الطوارئ). ويُوضِّح الشكل ٤-٨ مصفوفة الدقة لمشكلة تصنيف من فئتين؛ إذ تمثل الأرقام على طول القطر من اليســار العلوي إلى اليمين السفلي القرارات الصحيحة، والأرقام خارج هذا القطر تمثل الأخطاء.

الفئة المتوقعة	الفئة المشاهدة\الحقيقية: إيجابي	الفئة المشاهدة\الحقيقية: سلبي
إيجابي	العدد الإيجابي الحقيقي (TP)	العدد الإيجابي الخاطيء (FP)
سلبي	العدد السلبي الخاطيء (FN)	العدد السلبي الحقيقي (TN)

شكل ٤-٨: مصفوفة Confusion بسيطة لجدولة نتائج تصنيف فئتين

ويقدِّم الجدول ٤-١ معادلات مقاييس الدقة الشائعة لنماذج التصنيف.

جدول ٤-١: مقاييس الدِّقة الشائعة لنماذج التصنيف

القياس	الوصف
الدِّقة = (TP+TN/TP) + (TN+FPFN)	نســبة الحــالات المصنَّفــة بشــكلٍ صحيــحٍ (إيجابيات وسلبيات) مقسومةً على إجمالي عدد الحالات.
معدل موجب حقيقي = TP/ (TP+FN)	(a.k.a. Sensitivity) نســبة الإيجابيات المصنَّفة بشكلٍ صحيح مقســومةً على إجمالي العدد الموجب (أيْ: معدل الضرب أو الاسترجاع)
معدل سالب حقيقي = TN/ (TN+FP)	(a.k.a. Specificity) نســبة الســلبيات المصنَّفة بشكلٍ صحيح مقسومة على العدد الإجمالي السلبي (أيْ: معدل الإنذار الخطأ)

القياس	الوصف
الإحكام = (TP/TP)+FP	نسبة الإيجابيات المصنَّفة بشكلٍ صحيح مقسومةً على مجموع الإيجابيات المصنَّفة بشكلٍ صحيح والإيجابيات التي تمَّ تصنيفها بشكلٍ غير صحيح
الاستدعاء = (TP/TP)+FN	نسبة الإيجابيات المصنَّفة بشكلٍ صحيحٍ مقسومةً على مجموع الإيجابيات المصنَّفة بشكل صحيح والسَّلبيات غير المصنفة بشكل صحيح

وعندما لا تكون مشكلة التصنيف ثنائيةً؛ تزداد مصفوفة الارتباك (مصفوفة مربعة بحجم العددِ الفريد لعناوين الفئات)، وتُصبِح مقاييس الدِّقة مُحدَّدة بمعدلات دقة الفئة ودقة التصنيف الإجمالية.

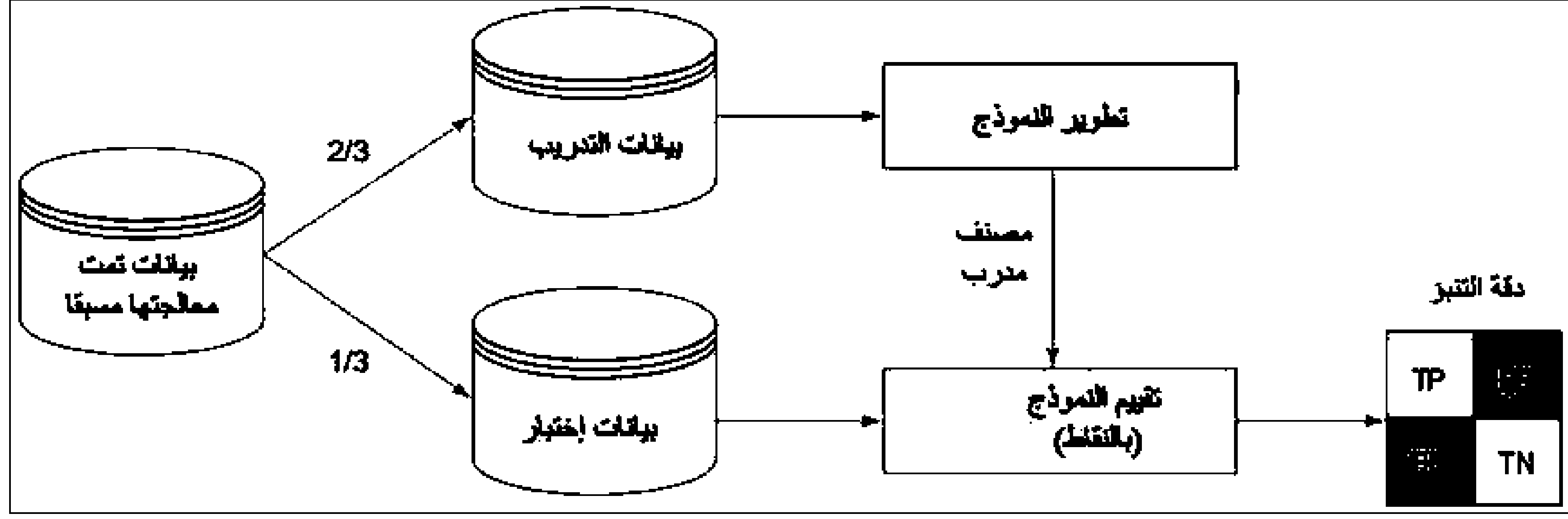
$$(\textit{True Classification Rate})_i = \frac{(\textit{True Classification})}{\sum_{i=1}^{n}(\textit{False Classification})}$$

$$(\textit{Overall Classifier Accuracy})_i = \frac{\sum_{i=1}^{n}(\textit{True Classification})_i}{\textit{Total Number of Cases}}$$

يُعَدُّ تقديرُ دقة نموذج تصنيف (أو مصنِّف) الناجم عن تعلُّم الخوارزمية تحت الإشراف أمرًا مهمًّا لسببين: أولًا: يمكِن استخدامه لتقدير مدى دقة التنبؤ المستقبلية، والتي قد تشير إلى مستوى الثقة الذي يجبُ أن يكون لدى المرء في مخرجات المصنف في نظام التنبؤ. وثانيًا: يمكن استخدامه لاختيار مصنف من بين مجموعة مصنَّفات (تحديد نموذج التصنيف «الأفضل» بين مجموعة مُدرَّبة).و فيما يلي بعض منهجيات التقدير الأكثر شيوعًا المُستَخدَمة لنماذج التنقيب في البيانات بطريقة التصنيف.

الانقسام البسيط: يُقسِّم الانقسام البسيط (أو إجراء تقييم أو اختبار للعيِّنة) البيانات إلى مجموعتين فرعيتين حصريتين متبادلتين وتُعرفان بمجموعة التدريب ومجموعة الاختبار (أو مجموعة التمرين)، ومن الشائع تعيينُ ثلثي البيانات كمجموعة تدريب من قِبل المُحفِّز (مُنشئ النموذج)، ومن ثم يتمُّ اختبار المصنِّف المُدمَج على مجموعة الاختبار. ويحدث استثناءٌ لهذه القاعدة عندما يكون المصنف عبارة عن شبكة عصبية اصطناعية. وفي هذه الحالة؛ يتم تقسيم

البيانات إلى ثلاث مجموعات فرعية اســتثنائية: التدريــب، والتحقق من الصحة، والاختبار. كما يتمُّ اســتخدام مجموعة التحقق من الصحة في أثناء بناء النموذج لمنع التثبيت. ويوضِّح الشــكل ٤-٩ منهجية الانقسام البسيط.



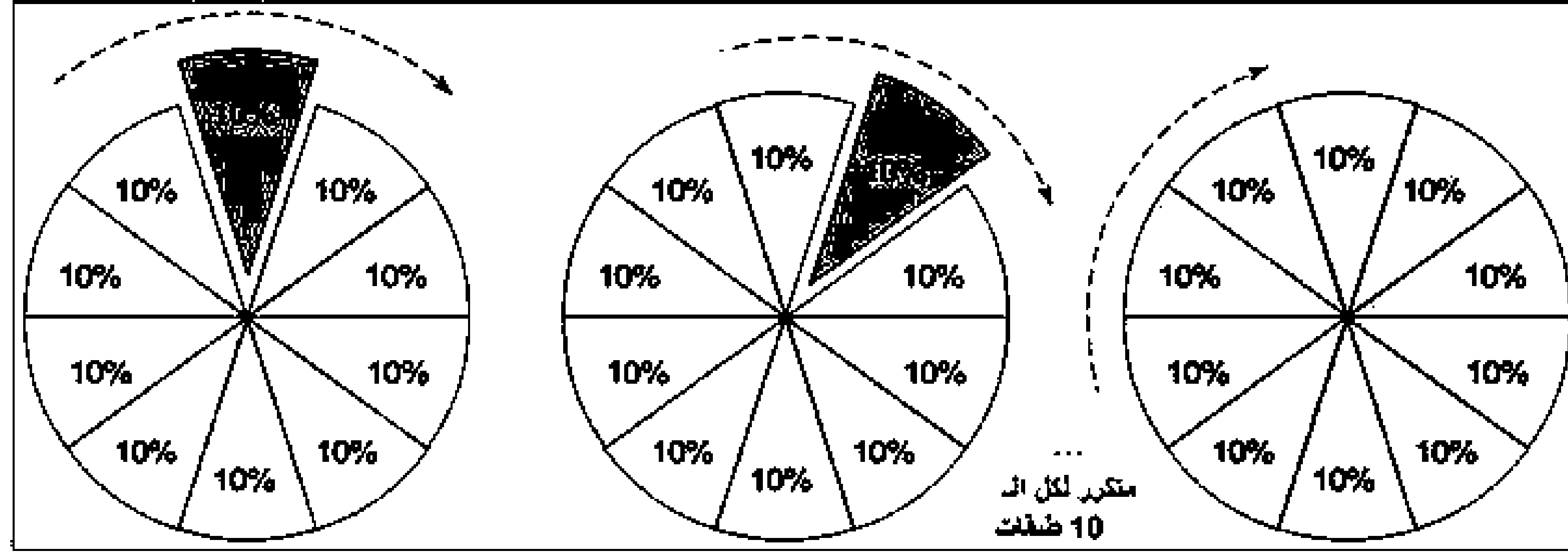
شكل ٤-٩: تقسيم البيانات العشوائية البسيطة

ويتمثل النقدُ الأساسي لهذه الطريقة في افتراض أن البيانات في المجموعتين الفرعيتين هي من نفس النوع (أيْ: لها نفس الخصائص بالضبط). ونظرًا لأن هذا التقســيم عشــوائي بسيط؛ فقد لا يكــون هذا الافتراض صحيحًا. ولتحســين هذا الوَضْع؛ فإنه يتمُّ اقتراح أخــذ العيِّنات الطبقية؛ إذ تصبح الطبقات هي متغيرُ المخرجات، وعلى الرغم من أنَّ هذا يُعَدُّ تحسُّنًا على الانقسام البسيط، غير أنه لا يزال لديه تحيُّز مرتبط بالتجزئة العشوائية المفردة.

التحقق باســتخدام س - أجزاء (K-flod): وتُســتَخدَم هذه المنهجيــة لتقليل التحيُّز المرتبط بالمعاينة العشــوائية لعيِّنات بيانات التدريب والعزل في مقارنة الدقة التنبؤية لطريقتين أو أكثر. وفي هذه المنهجية، والتي يُطلق عليها أيضًا تقدير الدوران، يتمُّ تقسيم مجموعة البيانات الكاملة بشــكلٍ عشــوائي إلى مجموعات فرعية حصرية متبادلة متســاوية الحجم تقريبًا؛ إذ يتمُّ تدريب نموذج التصنيف واختباره عدد (س) من المرات. وفي كلِّ مرة يتمُّ تدريبه على جميع أجزائه؛ ولكن مــرة واحدة ثم يتمُّ اختبارها على الجزء المنفرد المتبقي. ويتمُّ حســاب تقدير هذه المنهجية من الدقة الكلية للنموذج ببساطة بمتوسط مقاييس الدقة الفردية، كما هو مُوضَّح في المعادلة التالية:

$$CVA = \frac{1}{k}\sum_{i=1}^{k} A_i$$

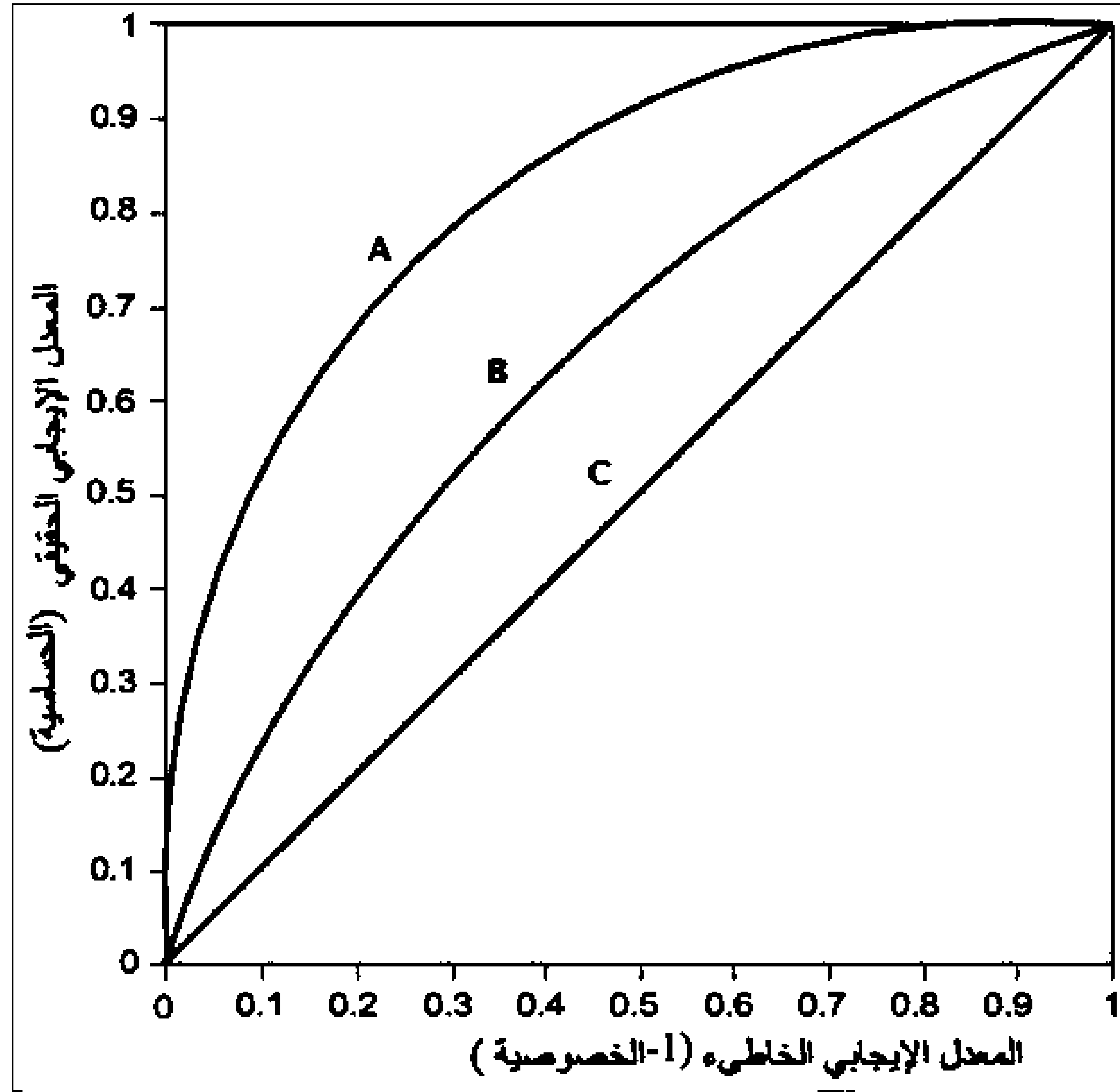
إذ يشيرُ CVA إلى دقة التحقق المتقاطع، و(س) هو عدد الأجزاء المُستَخدَمة، و(A) هو مقياسُ الدِّقة (أيْ: مُعدَّل الضرب، والحساسـية، والخصوصية) لكلِّ جزء. يعرض الشكل ٤-١٠ رسمًا بيانيًّا للتصديق المتقاطع k-fold؛ إذ تمَّ تعيين س على ١٠.



شكل ٤-١٠: تصوير بياني لاختبار الصلاحية المتقاطع من عدد K من الطبقات

منهجيات إضافية لتقييم التصنيف وتتضمَّن منهجيات التقييم الشائعة الأخرى ما يلي:

- **الإبقــاء على واحد Leave-one-out:** وهذه الطريقة تشـبه التحقُّق المتقاطع من K-fold؛ إذ تأخذ K قيمة ١؛ بمعنى أنه يتمُّ استخدام كلِّ نقطة بيانات للاختبار مرة واحدة على العديد من النماذج المتقدِّمة؛ إذ يُوجَد عددٌ من نقاط البيانات. وهذه الطريقة مُستَهلِكة للوقت؛ غير أنها في بعض الأحيان تُعَدُّ خيارًا قابلًا للتطبيق بالنسبة لمجموعات البيانات الصغيرة.
- **Bootstrapping:** ومــع هــذه الطريقة يتمُّ أخذ عينات عــددٍ ثابتٍ من النماذج من البيانات الأصلية (مع الاستبدال) للتدريب، وتُستَخدَم بقية مجموعة البيانات للاختبار. ويتمُّ تكرارُ هذه العملية عِدة مرات حسب الرغبة.
- **Jackknifing:** وعلى الرغم من تشابه هذه الطريقة مع منهجية Leave-one-out، مع احتساب التحصيل؛ فإنه يتمُّ حساب الدقة من خلال ترك عينة واحدة عند كلِّ تكرار لعملية التقدير.
- **المنطقــة تحت منحنى ROC:** إن المنطقة تحت منحنى ROC، هي تقنية تقييم رســومية؛ إذ يتــمُّ تخطيط المعدَّل الإيجابي الحقيقي على المحور Y، ويتمُّ رَسْــم المعدل الإيجابي الخاص على المحور X. وتحدد المنطقة الواقعة تحت منحنى ROC مقياس دقة المصنف؛ إذ تشــيرُ القيمة ١ إلى المصنف المثالي، في حين أنَّ ٠,٥ لا تشــير إلى أكثر من المصادفة العشــوائية؛ وفي الواقع؛ فإن القيم تتراوح بين الحالتين المتطرفتين. فمثلًا في الشكل ٤-١١، يحتوي A على أداء تصنيف أفضل من B، في حين أن C لا يكون أفضل من الفرصة العشوائية كرمي العملة.



شكل ٤-١١: منحنى عينة ROC

تقنيات التصنيف تستخدم عدداً من التقنيات (أو الخوارزميات)؛ لنمذجة التصنيف، ومنها:

- **تحليل شــجرة القرار:** يُمكِن القول بأن تحليل شجرة القرار (تقنية تعلُّم الآلة)، هو أكثر تقنيات التصنيف شــيوعًا في مجال التنقيب في البيانات. وســيرد وَصْفٌ مُفصَّلٌ لهذه التقنية في القسم التالي من هذا الفصل.
- **تحليل إحصائي:** كانت التقنيات الإحصائية عبارة عن الخوارزميات التصنيفية الأساســية، لعِدَّة ســنوات؛ حتى ظهــرت تقنيات تعلُّــم الآلة. وتتضَمَّــن تقنيات التصنيف الإحصــائي الانحدار اللوجيســتي والتحليل المميِّــز، وكلاهما من الافتراضات التي تفيد بــأن العلاقات بين المتغيرات المدخلة والمخرجات تكون بطبيعتها خطيةً، والبيانات موزَّعة بشكلٍ طبيعي، والمتغيرات ليست مترابطةً ومســتقلةً بعضهــا عن بعض. وقد أدَّت الطبيعة المشــكوك فيها لهذه الافتراضات إلى التحوُّل نحو تقنيات تعلُّم الآلة.
- **الشــبكات العصبية:** وهذه التقنية هي الأكثر شــيوعًا من بين تقنيات تعلُّم الآلة، والتي يُمكِن استخدامها لمشكلات نموذج التصنيف.

- **الاستدلال المبني على حالة:** يستخدم هذا النَّهج حالاتٍ تاريخيةً؛ للتعرُّف على القواسم المشتركة لتعيين حالة جديدة في الفئة الأكثر احتمالًا.
- **المُصنَّفات البيزية Bayesian classifiers:** يَستخدم هذا النهج نظريةَ الاحتمال لبناء نماذج تصنيف تستندُ إلى الأحداث السابقة القادرة على وَضْع مثيلٍ جديدٍ في فئة (أو سلسلة) أكثر احتمالًا.
- **الخوارزميات الجينية:** ويُقصَد بها استخدامُ تشابه التطوُّر الطبيعي لبناء آليات مستندة إلى البحث لتصنيف عيِّنات البيانات.
- **المجموعات الخام:** هذه الطريقة تأخذ بعين الاعتبار العضوية الجزئية لعلامات الفئات لفئات مُحدَّدة مُسْبقًا في نماذج البناء (تجميع القواعد) لمشكلات التصنيف.

والحقيقة أنَّ الوصف الكامل لجميع تقنيات التصنيف خارج نطاق هذا الكتاب؛ وبالتالي فإننا سوف نتناول هنا العديد من التقنيات الأكثر انتشارًا فقط.

مجموعة نماذج لأفضل التحليلات التنبؤية:

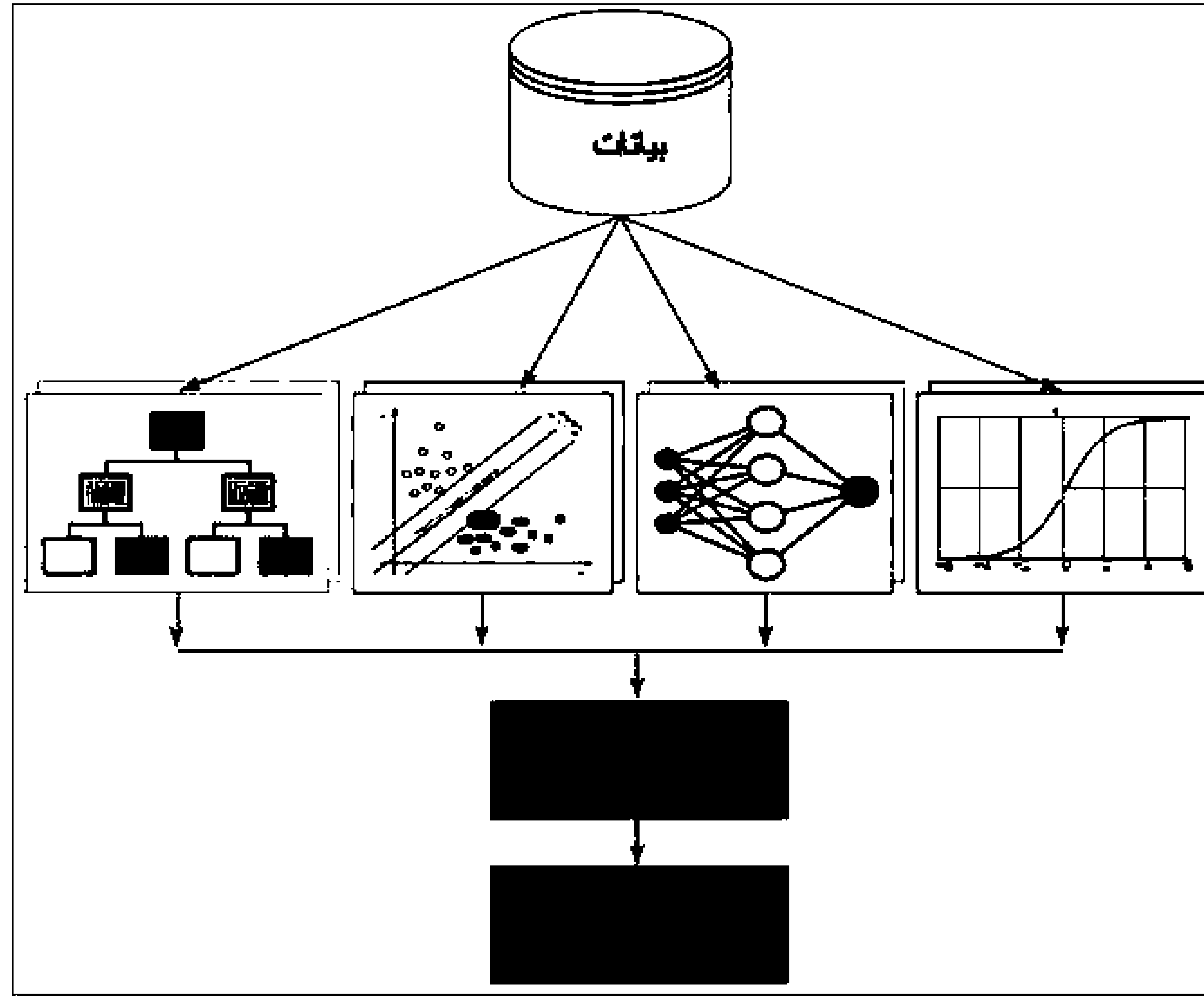
إنَّ عملية إنشاء مجموعات هي في جوهرها عمليةٌ ذكية لجمع المعلومات (التوقعات أو التنبؤات) التي تمَّ إنشاؤها وتوفيرها من قِبل اثنين أو أكثر من مصادر المعلومات (أيْ: نماذج التنبؤ). وعلى الرغم من وجود جدلٍ مستمرٍ حول مستوى تطوُّر أساليب المجموعة المُستخدَمة؛ فإن هناك إجماعاً عامًّا على أن نماذج المجموعات تُنتِج معلوماتٍ أكثر قوةً وموثوقيةً لقرارات العمل (Seni & Elder, 2010). أي إنَّ الجمع بين التنبؤات يُمكِن أن (وغالبًا ما) يؤدي إلى تحسين الدقة وقوة نتائج المعلومات، مع الحَدِّ من الشك والتحيُّز المرتبطين بالنماذج الفردية.

وكما نعلم جميعًا أنه لا يوجد في نماذج التنقيب في البيانات والتنبؤ بها نموذجٌ يُقال عنه أنه «أفضل نموذج» مقبول عالميًّا يناسب أيَّ مشكلة؛ إذ يعتمدُ أفضل نموذج على السيناريو الذي يجري تحليله ومجموعة البيانات المُستخدَمة؛ وهذا النموذج لا يُمكِن الحصول عليه إلا من خلال التجربة والخطأ على نِطاق واسع (وهذا لا يحدث إلا في حالة توافر الوقت والموارد). ومثلما لا يُوجَد نموذج واحد يمكن أن يُوصف بأنه الأفضل؛ فإنه لا يُوجَد أيضًا أفضل تنفيذ منفرد لأنواع النماذج المختلفة؛ فمثلًا أشجار القرار والشبكات العصبية وآلة المتجهات الداعمة لها أساليب بناء ومجموعات من العوامل المتغيِّرة تحتاجُ إلى «تحسين» أفضل النتائج الممكنة. ويقوم علماء البيانات بتطوير طرق جديدة لتحسين دقة وكفاءة نماذج التنبؤات في الوقت الحالي. وتتمثل إحدى الطرق المؤكدة للقيام بذلك في دَمْج مخرجات نماذج التنبؤ في نتيجة مركبة واحدة، وهي

مجموعة نموذجية. وتُعَدُّ الفِرق التي تستخدم مجموعة نماذج (Ensemble) - الفِرق التي فازت بالعديد من منافسات التنقيب في البيانات والتنبؤ بالبيانات في السنوات الأخيرة - هي الفِرق الفائزة (وللحصول على قائمةٍ بمسابقات التحليلات والفرص التحليلية الحديثة انظر: kaggle.org).

ويُمكِن تصنيف نماذج المجموعات إلى متجانسة أو غير متجانسة (Abbott 2014, P. 307). وكما يُوحِي الاسم؛ فإن مجموعات النماذج المتجانسة تجمع بين نتائج نموذجين أو أكثر من نفس النوع من النماذج، مثل أشجار القرار. وفي الواقع؛ فإن تطوير أغلب مجموعات النماذج المتجانسة يتمُّ باستخدام مزيجٍ من هياكل شجرة القرار. وتُعَدُّ فِئتا التعبئة والتعزيز الفئتين الشائعتين من مجموعات شجرة القرار. ويُطلق على المثال المعترَف به بشكلٍ كبيرٍ والذي حقق نجاحًا كبيرًا في مجموعات شجرة القرار من نوع مجموعات الغابة العشوائية Random Forest - بدلًا من بناء شجرة كبيرة؛ إذ تتطوَّر الغابة العشوائية Random Forest كغابة من العديد من الأشجار الصغيرة. وخير مثالٍ لدعم مجموعات شجرة القرار هو AdaBoosting (وهو اسم قصير لـ «التعزيز التكيُّفي») - وهو الخوارزمية التي تعمل على تغيير الوزن (أيْ: الأهمية أو المساهمة) المخصص لكلِّ عينة بيانات في كل تكرار في عملية التعلُّم، والذي يستند إلى نتائج التصنيف غير الصحيح؛ بحيث يتمُّ تحسينُ دقة المصنف / المتنبئ لجميع تصنيفات الفصل.

أمَّا مجموعات النماذج غير المتجانسة؛ فإنها - كما يوحي اسمها - تجمع بين نوعين مختلفين أو أكثر من النماذج، مثل: أشجار القرار، والشبكات العصبية الاصطناعية، والانحدار اللوجستي، وآلة المتجهات الداعمة. والجدير بالذكر أن استخدام النماذج التي تختلف جوهريًّا بعضها عن بعض يُعَدُّ أحد عوامل النجاح الرئيسة في نمذجة المجموعات؛ إذ تتناول البيانات من منظور مختلف. ونظرًا لأن الطريقة التي تجمع بين نتائج نماذج مختلفة لأنواع مختلفة من النماذج؛ فإن المجموعة النموذجية غير المتجانسة تُسمَّى أيضًا نماذج الدَّمج المعلوماتي (Delen وSharda, 2010). وفي عملية دَمْج نتائج نماذج متعددة، سواء أكانت تصويتاً بسيطاً (كلُّ نموذج يُسهم بالقدر نفسه، صوت واحد) أو مزيج تركيبي من التصويت (يُسهم كلُّ نموذج بناءً على دقة التنبؤ به - النماذج الأكثر دقةً لها قيمة ووزن أعلى) ممكن استخدامه. وبغض النظر عن طريقة التوليفة؛ فقد ثبُت أن المجموعات تمثل إضافةً لا تُقدَّر بثمن لأيِّ مشروع للتنقيب في البيانات والنمذجة التنبؤية. وعلى الرغم من أنَّ النمذجة المُعزَّزة تعمل على تحسين الدقة والقوة، غير أن لها جانباً سلبيًّا؛ إذ إنها تزيدُ أيضًا من تعقيد النموذج، ومن ثم عدم القدرة على تفسيرها (أيْ: الشفافية). ويُوضِّح الشكل ٤-١٢ رسمًا بيانيًّا لمجموعة نموذج غير متجانس.



شكل ٤-١٢: توضيح بياني لمجموعة غير متجانسة

وقبل أن نوضِّح تفاصيل أشـجار القرار؛ فإننا نحتاجُ إلى مناقشـة بعض المصطلحات البسيطة. أولًا: تشـمل أشـجار القرار العديدَ من متغيرات المدخلات التي قد يكون لها تأثيرٌ على تصنيف الأنماط المختلفة. وعادةً ما تُسـمَّى هذه المتغيرات المُدخَلة بالسِّمات. فمثلًا إذا أردنا بناء نموذجٍ لتصنيف مخاطر القروض على أساس اثنين فقط من الخصائص - الدخل والتصنيف الائتماني - فإنَّ هاتين الخاصيتين سـتكونان هما الصفات، والمخرجات الناتجة ستكون هي الفئة (والتي قد تكون منخفضةً أو متوسطةً أو عالية المخاطر). ثانياً: تتكون الشجرة من مجموعة من الفروع والعقد. ويمثل الفرع نتيجة اختبار لتصنيف نمط (على أسـاس اختبار) باسـتخدام إحدى السِّمات. وتمثل عقـدة الورقة في النهاية اختيار الفصل النهائي للنمط (سلسـلة من الفـروع من عقدة الجذر إلى عقدة الورقة، والتي يمكن تمثيلها كإشارة if-then معقدة).

وتكمن الفكرة الأساسـية وراء شـجرة القرار في أنها تقسِّم بشـكلٍ متكرِّر مجموعة التدريب حتى يتكون كلُّ قسم بشكلٍ كاملٍ أو بشكلٍ أساسي من أمثلة من فصل واحد. وتحتوي كلُّ عقدة غير ورقية من الشـجرة على نقطة انقسـام، وهي عبارة عن اختبار لواحد أو أكثر من السـمات وتُحدِّد كيفية تقسيم البيانات بشكلٍ أكبر. وبشكلٍ عام؛ فإن الخوارزميات الخاصة بشجرة القرار تقوم بإنشاء شجرة مبدئية من بيانات التدريب؛ بحيث تكون كلُّ عقدة عبارة عن أوراق نقية، ثم تقوم بتقليم الشجرة لزيادة تعميمها، وبالتالي دقة التنبؤ على بيانات الاختبار.

وفي مرحلة النموِّ؛ يتمُّ بناء الشــجرة عن طريق تقسيم البيانات بشكلٍ مُتكرِّر حتى تصبح كل وحدة نقية (أيْ: تحتوي على أعضاء من نفس الفئة) أو صغيرة نســبيًّا. وفكرتها الأساســية تعتمد على طرح الأســئلة والتي تقوم الإجابات عنها بتوفير معظم المعلومات، على غِرار ما قد نفعله إذا لعبنا لعبة «عشرون سؤالًا».

ويعتمد الانقسامُ المستخدَم لتقسيم البيانات على نوع السِّمة المُستَخدَمة في التقسيم. وبالنسبة إلى السِّــمة المســتمرة A، تكون التقســيمات من قيمة النموذج X > (A)؛ إذ تمثل X قيمة الانقسام «المثلى» لـ A. فعلى سبيل المثال يمكن أن يكون التقسيم المبني على الدخل هو «الدخل ٥٠٠٠٠ >». وبالنسبة للسِّمة الفئوية A؛ فإن الانقسامات التي هي من قيمة النموذج (A) تنتمي إلى X؛ حيث X هي مجموعة فرعية من A. فمثلًا يمكن أن يكون الانقسام على أساس الجنس: «ذكر مقابل أنثى».

أمَّا الخوارزمية العامة لبناء شجرة القرار؛ فهي كما يلي:

١- إنشاء عقدة جذرية وتعيين جميع بيانات التدريب إليها.

٢- تحديد أفضل سمة تجزئة.

٣- إضافة فرع إلى عقدة الجذر لكلِّ قيمة من الانقسام؛ بحيث يتمُّ تقسيم البيانات إلى مجموعات فرعية متبادلة (nonoverlapping) على طول خطوط التقسيم المُحدَّد والانتقال إلى الفروع.

٤- تكرير الخطوتين ٢ و٣ لكلِّ عقدة موجودة في كلِّ ورقة حتى الوصول إلى معايير التوقف (فمثلًا يتمُّ التحكم في العقدة بتسمية فئة واحدة).

وقــد تم اقــتراحُ العديد من الخوارزميات المختلفة لإنشــاء أشــجار القــرار. وتختلف هذه الخوارزميــات بشــكلٍ أســاسي من حيــث الطريقة التــي تُحدّد بهــا خاصية التجزئــة (وقيمها المقسومة)، وترتيب تقسيم السمات (تقسيم نفس السمة مرة واحدة فقط أو عدة مرات)، وعدد الانقســامات عند كلِّ عقدة (ثنائي مقابل ثلاثي)، ومعايير التوقف، وتقليم الشجرة (ما قبل الحرق postpruning). ومن أكثر الخوارزميات شــهرةً نظام ID3 (متبوعة بـ C4.5 وC5 كنسخ معدلة مــن ID3) من تعلم الآلة، والتصنيف، بالإضافة إلى أشــجار الانحدار (CART) من الإحصائيات، وكاشف التفاعل الأوتوماتيكي مربع كاي chi-squared (CHAID) من التعرُّف على الأنماط.

وعند إنشــاء شجرة قرار؛ يكون الهدف في كلِّ عقدة هو تحديدُ السِّمة ونقطة الانقسام لهذه السمة التي تقسِّم سجلات التدريب على أفضل وجه لتنقية تمثيل الفئة في تلك العقدة. ولتقييم ســلامة الانقســام؛ فقد تمَّ اقتراحُ بعض مؤشرات التقسيم. وقد وُجِدَ أن أكثر مؤشرين شيوعًا هما

مؤشر جيني Gini index وكسب المعلومات information gain؛ حيث يُستَخدَم مؤشر جيني Gini index في خوارزميات CART وSPRINT (قابلة للتحجيم بالتعقب الذاتي للأشجار). كما يتمُّ استخدام إصدارات من كسب المعلومات في ID3 (والإصدارات الأحدث، C4.5 وC5).

وقد استُخدِم مؤشر جيني Gini index في الاقتصاد لقياس تنوع السكان. كما يمكن استخدام نفس المفهوم لتحديد درجة نقاء فئة معينة كنتيجة لقرار التفرُّع على طول سمة أو متغير معين. وأفضل انقسام هو الذي يزيدُ من نقاء المجموعات الناتجة عن الانقسام المقترح. وفيما يلي نتناول بإيجاز حساباً بسيطاً لمؤشر جيني.

إذا كانت مجموعة البيانات S تحتوي على أمثلة من فئات n؛ فسيتمُّ تعريف مؤشر جيني Gini index على أنه

$$gini(S) = 1 - \sum_{j=1}^{n} p_j^2$$

حيث Pj هو تردد نسبي للفئة j في S. فإذا كانت مجموعة البيانات S مقسمةً إلى مجموعتين فرعيتين، S1 وS2، بأحجام N1 وN2، على التوالي؛ فإنَّ مؤشر جيني Gini index للبيانات المقسَّمة يحتوي على أمثلة من فئات n، ويتم تعريف مؤشر جيني Gini index على أنه

$$gini_{split}(S) = \frac{N_1}{N} gini(S_1) + \frac{N_2}{N} gini(S_2)$$

حيث يتمُّ اختيار مجموعة السمة / الانقسام التي توفر أصغر ginisplit (S)؛ من أجل تقسيم العقدة. وفي مثل هذا الشأن؛ فإنه يجب سرد جميع نقاط التقسيم الممكنة لكلِّ صفة.

أمَّا كَسْب المعلومات؛ فهو آلية التقسيم المُستَخدَمة في ID3، والتي ربما تكون الخوارزمية الأكثر شهرةً والمعروفة باسم شجرة القرار. وقد تمَّ تطويرها من قِبل Ross Quinlan في ١٩٨٦م، ومنذ ذلك الحين؛ فقد قام بتطوير هذه الخوارزمية ضِمن الخوارزميات C4.5 وC5. وتكمن الفكرة الأساسية لـ ID3 (ومتغيراتها) في استخدام مفهومٍ يُسمَّى الإنتروبيا entropy بدلًا من مؤشر جيني Gini index؛ حيث يقيس الانتروبيا مقدار الشك أو العشوائية في مجموعة البيانات. فإذا كانت جميع البيانات في مجموعةٍ فرعيةٍ تنتمي إلى فئة واحدة فقط؛ فلا يُوجَد أيُّ شك أو

عشوائية في هذه المجموعة من البيانات، وبالتالي؛ فإن الإنتروبيا entropy صفر. والهدف من هذا النهج؛ هو بناء الأشـجار الفرعية بحيث تكون الإنتروبيا entropy لكل شـجرة فرعية نهائية صفر (أو قريبة من الصفر). وفيما يلي نتناول أيضًا حساب كسب المعلومات.

وعـلى افـتراض أنَّ هنـاك فئتـين، P (موجبـة) وn (سـالبة). دع مجموعـة الأمثلـة S تحتـوي عـلى عـدد p من فئـة P وعـدد n مـن الفئة N. فـإن كميـة المعلومـات المطلوبة لتحديـد مـا إذا كان المثـال التعسُّـفي في S ينتمـي إلى P أو N يتـمُّ تعريفهـا عـلى أنهـا

$$I(p, n) = -\frac{p}{p+n}\log_2\frac{p}{p+n} \quad \frac{n}{p+n}\log_2\frac{n}{p+n}$$

وبفرض اسـتخدام السمة A، فإنه سيتم تقسـيم المجموعة S إلى مجموعات {S1, S2, …, Sv}. فـإذا احتوى Si عـلى أمثلة pi من P وأمثلـة ni من N، فإن الإنتروبيـا entropy، أو المعلومات المتوقعة اللازمة لتصنيف الكائنات في كل الأشجار الفرعية، Si، هي

$$E(A) = \sum_{i=1}^{n} \frac{p_i + n_i}{p + n} I(p_i, n_i)$$

ومـن ثَـمّ؛ فـإن المعلومات التي سـيتم اكتسـابها من خـلال التفرُّع على السِّـمة A سـتكون Gain (A) = I (p,n) - E (A).

حيث يتمُّ تكرار هذه الحسـابات لكلِّ سـمة، كما يتمُّ تحديدُ تلك التي لها كسب المعلومات العالي كسـمة التجزئة. وتتشابه الأفكار الأساسـية وراء مؤشرات التقسيم هذه بعضها مع بعض، غير أن تفاصيل الخوارزمية المُحدَّدة تختلف. ويمكن العثورُ على تعريف تفصيلي لخوارزمية ID3 وآلية تقسيمها في Quinlan (1986).

وتوضِّح الحالة العملية ٤-٥ مدى أهمية المكاسـب المتحصَّل عليها إذا تمَّ اسـتخدام تقنيات التنقيب في البيانات الصحيحة لمشكلة عمل مُحدَّدة بشكلٍ جيد.

حالة عملية ٤-٥

تستخدم شركة Influence Health التحليلات التنبؤية المتقدِّمة للتركيز على العوامل التي تؤثر حقًّا في قرارات الرِّعاية الصحية الخاصة بالأشخاص

تقدِّم شركة Influence Health منصة تفاعل وتفعيل المستهلك الرقمي الوحيد في قطاع الرعاية الصحية؛ حيث تُمكِّن منصة Influence Health مقدِّمي الخدمة وأصحاب العمل والدافعين من التأثير إيجابًا على صُنْع القرار والسلوكيات الصحية للمستهلك بما يتجاوز إعداد الرعاية البدنية من خلال المشاركة المتعددة الشخصية والتفاعلية. وقد ساعدت الشركة منذ عام ١٩٩٦م - ومقرها في برمنغهام بولاية ألاباما - أكثر من ١١٠٠ منظمة مزوِّدة في التأثير على المستهلكين بطريقة تحوُّلية لتوفير التكاليف بالإضافة إلى الجودة.

إن الرعاية الصحية هي عملٌ شخصي؛ إذ تختلف احتياجات كلِّ مريض وتتطلب استجابة فردية، ومن ناحيةٍ أخرى - فإنه مع استمرار ارتفاع تكاليف توفير خدمات الرعاية الصحية- تحتاج المستشفيات والنظم الصحية بشكلٍ متزايد إلى الاستفادة من وفورات الحجم من خلال توفير الطعام لأكبر عددٍ من السكان؛ إذ يصبح التحَدِّي عندئذٍ هو توفير نهج شخصي في أثناء العمل على نطاق واسع. كما تتخصَّص شركة Influence Health في مساعدة عملاء قطاع الرعاية الصحية على حلِّ هذا التحَدِّي من خلال التعرُّف على مرضاهم الحاليين والمحتملين بشكلٍ أفضل واستهداف كلِّ فرد بالخدمات الصحية المناسبة في الوقت المناسب. وتسمح تقنية التحليلات التنبؤية المتقدِّمة من IBM لشركة Influence Health بمساعدة عملائها على اكتشاف العوامل التي لها أكبر تأثيرٍ على قرارات الرعاية الصحية للمرضى. ومن خلال تقييم ميل مئات الملايين من العملاء المتوقعين للحصول على خدمات رعاية صحية خاصة؛ فإن Influence Health قادرةٌ على زيادة الإيرادات ومعدلات الاستجابة لحملات الرعاية الصحية، وتحسين النتائج لعملائها ومرضاهم على حَدٍّ سواء.

استهداف المستهلك الذكي:

أصبحت صناعة الرعاية الصحية اليوم أكثر تنافسيةً من أيِّ وقتٍ مضى؛ إذ إنَّ انخفاض استخدام خدمات المنظمة؛ يؤدي إلى انخفاض أرباحها أيضًا؛ إذ يميل المستهلكون الآن إلى اختيار أفضل مُقدِّمي الرعاية الصحية، بدلًا من مجرد البحث عن أقرب مستشفى أو عيادة، وبالتوازي مع الجهود المشتركة في الصناعات الأخرى؛ فإنه يجب على منظمات

الرعاية الصحية بذل المزيد من الجهد لتسـويق نفسـها بشـكلٍ فعّال لكلٍّ من المرضى الحاليين والمحتملين وبناء التواصل والولاء على المدى الطويل.

إن مفاتيـح النجاح في تسـويق الرعاية الصحية؛ هي التوقيت المناسـب والأهمية. فإذا كنت تستطيع التنبؤ بنوع الخدمات الصحية التي قد يحتاجها الفرد المحتمل؛ فإنه يمكنك المشاركة والتأثيرُ عليه بشكلٍ أكثر فاعليةً في العناية بالصحة.

يُوضِّح Venky Ravirala رئيس قسـم التحليلات في شركة Influence Health، أن: «مؤسسـات الرِّعاية الصحية تخاطر بفقدان انتباه الأشـخاص إذا أوصلوا إليهم رسـائل غير ذات صلة. ونحن نساعد عملاءنا على تجنب هذا الخطر من خلال استخدام التحليلات لتقسيم آفاقهم الحالية والمحتملة والسوق لهم بطريقة شخصية وملائمة أكثر بكثير».

تحليلات أسرع وأكثر مرونة:

ومع توسُّـع قاعدة عملائهـا؛ ازداد الحجم الإجـمالي للبيانات في أنظمـة التحليلات في Influence Health؛ ليشـمل أكثر من ١٩٥ مليون سجل للمرضى، مع سجل تفصيلي للمرض لعِـدَّة ملايـين من المـرضى. وقد علق Ravirala عـلى ذلك بقوله: «مع وجـود الكثير من البيانات للتحليل؛ أصبحت طريقة تسـجيل البيانات الحالية لدينا معقدةً للغاية وتسـتغرق وقتاً طويلاً. ولذلك فقد أردنا أن نكون قادرين على استخراج الأفكار بسرعة أكبر ودقة أكبر».

ومـن خلال الاسـتفادة مـن برمجيـات التحليـلات التنبؤية مـن IBM؛ أصبحت Influence Health الآن قـادرةً على تطوير نماذج تحسـب مـدى احتمالية أن يحتاج كلُّ مريض إلى خدمات معينة وتُعبِّر عن هذا الاحتمال كنسبة مئوية. ويعتمد التقسيم الجزئي والعديد من النماذج الخاصة بالأمـراض على البيانات الديموغرافية والاجتماعية والاقتصادية والجغرافية والسـلوكية وتاريخ الأمـراض وبيانات التعداد وبحث الجوانب المختلفة لاحتياجات كلِّ مريض من الرعاية الصحية المتوقعة.

ويضيـف Ravirala قائـلًا: «ويتيح لنا حـلُّ IBM الجمع بـين كلِّ هذه النماذج باسـتخدام تقنية المجموعة؛ مما يساعد على التغلُّب على قيود النماذج الفردية وتقديم نتائج أكثر دقةً؛ فهي تمنحنا المرونة لتطبيق تقنيات متعددة لحلِّ مشـكلةٍ ما والتوصُّل إلى أفضـل حَـلٍّ لها. كما أنها تعمـل على تنفيذ الكثير من عمليـات التحليلات تلقائيًّا؛ مما يُمكِّننا من الاسـتجابة لطلبات العملاء بشكلٍ أسرع من ذي قبل، وكثيرًا ما نوفر لهم مستوى أعمق من البصيرة فيما يتعلق بالقطاع المرضى».

فعلى ســبيل المثــال: قرَّرت Influence Health اكتشــاف كيفية انتشــار المرض وتنوع الخطر بين مجموعات مختلفة داخل عامة السكان. وذلك باستخدام تقنيات تحليل مجموعة معقدة للغاية، تمكَّن الفريق من اكتشــاف أمراض الاعتلال المشتركة الجديدة التي تعمل على تحسين القدرة على التنبؤ بالمخاطر لأكثر من ١٠٠ مرض شائع بنسبة تصل إلى ٨٠٠٪.

وهذا يســاعد على التفريق بثقةٍ تامةٍ بين المــرضى ذوي المخاطر العالية والمعرضين لمخاطر عالية للغاية؛ مما يســهِّل اســتهداف الحملات على المرضى والمحترفين الذين هم في أمــسِّ الحاجــة إليها. ومع وجود مثل هذه الأفــكار في متناول اليد؛ فإن Influence Health قادرةٌ على استخدام خبرتها في تسويق الرعاية الصحية لتقديم المشورة لعملائها حول أفضل السُّبل لتخصيص موارد التسويق.

وقــد أردف Ravirala قائــلًا: «يقوم عملاؤنا بعمل ضوابط كبيرة للميزانية على أســاس التوجيهــات التي نقدمها لهم». «ونحن نســاعدهم على زيادة تأثــير حملات لمرة واحدة - مثــل حملات ســوق التأمين الصحي عنــد بــدء Obamacare - بالإضافة إلى خططهم الإستراتيجية الطويلة المدى والاتصالات التسويقية المستمرة».

الوصول إلى الجمهور المناسب:

يســاعد تأثيرُ شركــة Influence Health على زيــادة الإيرادات وتحســين الصحة الســكانية؛ من خلال تمكين عملائها من اســتهداف أنشــطتهم التســويقية بشكلٍ أكثر فعاليــةً. ويُوضِّــح Ravirala: «وبفضل العمــل معنا؛ تمكَّن العملاء مــن تحقيق عائد اســتثماري يصل إلى ١٢-١ من خلال التسويق المُســتَهدَف بشكلٍ أفضل، لا يتعلق الأمر فقــط بالإيرادات: فمــن خلال ضمان إرســال معلومات الرعاية الصحيــة الحيوية إلى الأشــخاص الذين يحتاجون إليها، نحن نســاعد عملاءنا على تحســين مستويات الصحة العامة في المجتمعات التي تخدمهم».

وتســتمرُّ Influence Health في صقــل تقنيــات النمذجة، واكتســاب فهمٍ أعمق للســمات الحرجة التي تؤثر على قرارات الرعاية الصحية. فمع مجموعة أدوات تحليل مرنــة في متناول يدها؛ فإن الشركة مجهزةٌ بشــكلٍ جيدٍ للحفاظ على تحســين خدماتها للعمــلاء. ويذكر Ravirala أنه: «في المســتقبل؛ نريد أن نتعــرف على بيانات المريض والبيانات المرتقبة إلى المستوى التالي، ونحدد أنماط السلوك وندمج التحليل مع مكتبات تعلــم الآلــة. لقد أعطتنــا IBM SPSS بالفعل القدرة على تطبيــق ودمج العديد من

الموديلات دون كتابة خط واحد من التعليمات البرمجية. ونحن حريصون على الاستفادة من حلِّ IBM مع توسيع نطاق تحليلات الرِّعاية الصحية لدينا لدعم النتائج السريرية وخدمات إدارة صحة السكان».

ويختتم Venky Ravirala، رئيس قسم التحليلات في Influence Health قائلًا: «نحن نحقق تحليلات على نطاقٍ غير مسبوق. فاليوم؛ يمكننا تحليل ١٩٥ مليون سجل مع ٣٥ نموذج مختلف في أقل من يومين، وهي مهمةٌ لم تكن ممكنة لنا في الماضي».

أسئلة للمناقشة:

١- ماذا فعلت Influence Health؟

٢- ما هي التحديات والحلول المقترحة والنتائج التي تمَّ الحصول عليها؟

٣- كيف يمكن أن يساعد التنقيب في البيانات الشركات في صناعة الرعاية الصحية (بطرق أخرى غير المذكورة في هذه الحالة)؟

Source: Influence Health: Focusing on the factors that really influence people's healthcare decisions. IBM Case Study, www.presidion.com/case-study-influence-health.

التحليل التجميعي للتنقيب في البيانات:

يُعَدُّ التحليلُ التجميعي طريقةً أساسيةً للتنقيب في البيانات لتصنيف العناصر أو الأحداث أو المفاهيم إلى مجموعات مشتركة تُسمَّى مجموعات. وتُستخدَم هذه الطريقة بشكلٍ شائعٍ في علم الأحياء، والطب، وعلم الوراثة، وتحليل الشبكات الاجتماعية، وعلم الإنسان، وعلم الآثار، بالإضافة إلى علم الفلك، والتعرُّف على المفاهيم، وحتى في نظم إدارة المعلومات (MIS). ونظرًا لزيادة انتشار التنقيب في البيانات؛ فقد تمَّ تطبيق التقنيات الأساسية على الأعمال؛ خاصةً على التسويق. كما تمَّ استخدامُ التحليل التجميعي على نطاقٍ واسعٍ؛ للكشف عن الاحتيال (كلٌّ من الاحتيال على بطاقة الائتمان والتجارة الإلكترونية) وتجزئة السوق للعملاء في أنظمة CRM المعاصرة. ويستمر تطوير المزيد من التطبيقات في الأعمال؛ إذ يتمُّ التعرُّف على قوة تحليل المجموعات واستخدامها.

إن التحليل التجميعي هو أداةٌ لتحليل البيانات الاستكشافية لحلِّ مشكلات التصنيف. ويهدف إلى تصنيف الحالات (مثل: الأشخاص، والأشياء، والأحداث) إلى مجموعات أو مجموعات؛ بحيث تكون درجة الارتباط قويةً بين أعضاء نفس المجموعة وضعيفةً بين أعضاء المجموعات المختلفة. وتصف كلُّ مجموعة الفئة التي ينتمي إليها أعضاؤها. ومن الأمثلة الواضحة في تحليل الأبعاد

التجميعية؛ إنشـاء نطاقات درجات يُمكِن من خلالها تعيين درجات دراسـية لفئة الكلية. وهذا يشبه مشكلة التحليل التجميعي التي واجهتها وزارة الخزانة الأمريكية عند إنشاء أقواس ضريبية جديـدة في الثمانينيـات. مثال خيالي لحـدوث التجميع في كتب هاري بوتـر J. K. Rowling's Harry Potter books. يُحـدِّد Sorting Hat أي بيـت (مبنى) لتعيين طلاب السـنة الأولى في مدرسـة هوجورتس. مثالٌ آخر ينطوي على تحديـد كيفية جلوس الضيوف في حفل الزفاف. أمَّا فيما يتعلق بمعالجة البيانات؛ فإن أهمية التحليل التجميعي هي أنها قد تكشف عن مجموعات وهياكل في بيانات لم تكن واضحةً من قبل ولكنها معقولة ومفيدة عندما يتم العثور عليها.

ويمكن استخدامُ نتائج التحليل التجميعية في:

- تحديد نظام التصنيف (مثل: أنواع العملاء).
- اقتراح نماذج إحصائية لوصف السكان.
- ظهور قواعد لتعيين حالات جديدة للفصول لتحديدها واستهدافها وأغراض التشخيص.
- توفير مقاييس التعريف والحجم والتغيير في المفاهيم السابقة.
- البحث عن حالات نموذجية لتسمية وتمثيل الطبقات.
- تقليل حجم وتعقيد مساحة المشكلة للطرق الأخرى للتنقيب في البيانات.
- تحديد القِيم المتطرفة في مجالٍ معين (مثل: اكتشاف الأحداث النادرة).

تحديد العدد الأمثل للمجموعات: تتطلَّب الخوارزميات التجميعية عادةً تحديدَ عدد المجموعات المطلـوب العثور عليهـا. فإذا كان هذا الرقم غـير معروف من خلال المعرفة السـابقة؛ فإنه يجب اختياره بطريقة ما. ولسـوء الحظ؛ فإنه لا تُوجَد طريقة اختيارية لحسـاب ما يفترض أن يكون هذا الرقم. ولذلك فقد تمَّ اقتراحُ عِدة طرق إرشادية مختلفة. وفيما يلي نذكر أكثر هذه الطرق شيوعًا:

- انظر إلى النسبة المئوية للفارق الموضَّح كدالة لعدد المجموعات؛ أيْ: اختر عددًا من المجموعات؛ بحيـث لا يؤدي إضافة مجموعة أخـرى إلى إعطاء نماذج أفضل للبيانات. وعلى وجه التحديد، إذا تـمَّ توضيحُ النِّسـب المئوية للتباين في أحد الرسـوم البيانية بواسـطة المجموعات؛ فهناك نقطةٌ سـينخفض عندها الكسـب الهامشي (إعطاء زاوية في الرسـم البياني)؛ مما يشير إلى عدد المجموعات التي سيتمُّ اختيارها.
- اضبط عددَ المجموعات على (n / ٢) ٢ / ١، حيث n هو عدد نقاط البيانات.
- اسـتخدم معيار المعلومات AIC) Akaike)، وهو مقياسٌ للصلاحية الملائمة (على أساس مفهوم الانتروبيا entropy) لتحديد عدد المجموعات.

- اســتخدِم معيار المعلومــات Bayesian، وهو معيار اختيار النموذج (على أســاس الحَدِّ الأقصى لتقدير الاحتمال) لتحديد عدد المجموعات.

طرق التحليل: قد يعتمد تحليل المجموعات على واحدٍ أو أكثر من الطرق العامة التالية:

- الأساليب الإحصائية (بما في ذلك التسلسل الهرمي وغير الهرمي)، مثل: k-means أو k-modes.
- الشبكات العصبية (مع أسلوب بناء يُسمَّى خريطة التنظيم الذاتي).
- المنطق الضبابي (مثل: خوارزمية fuzzy C-means).
- الخوارزميات الجينية.

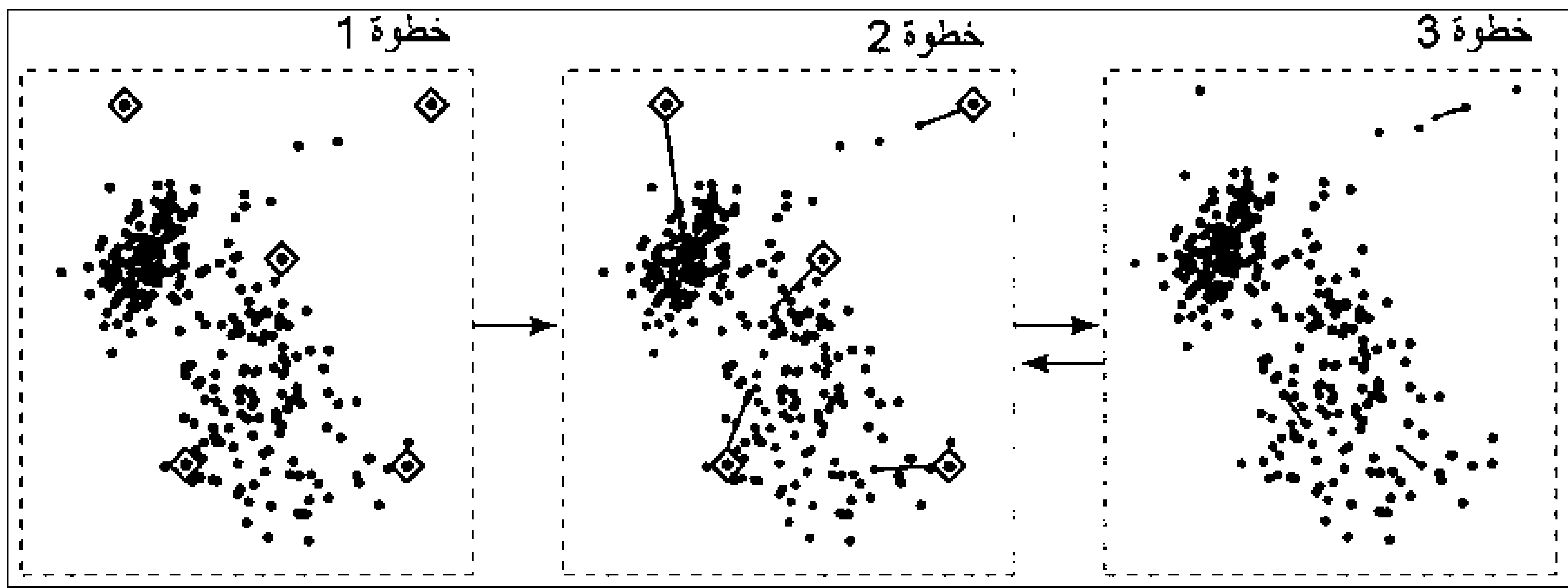
تعمل كلُّ طريقة من هذه الطرق بشكلٍ عام مع واحدة من فئتي الطريقة العامة:

- انقسام: مع الفصول المقسَّمة، تبدأ جميع العناصر في مجموعةٍ واحدة وتتفكك.
- تجميــع هرمــي: مــع الفئات التجمعيــة، تبدأ جميع العنــاصر في مجموعــات فردية، وتكون المجموعات مجمعة معًا.

وتتضمَّن معظم طرق التحليل التجميعية استخدام مقياس المسافة؛ لحساب التقارب بين أزواج العناصر. وتشــمل مقاييس المسافة المنتشرة المســافة التقليدية (أيْ: المسافة العادية بين نقطتين يمكن قياسها بالمسطرة) ومسافة مانهاتن (وتُسمَّى أيضًا المسافة المستقيمة، أو مسافة التاكسي بين نقطتين). وفي كثيرٍ من الأحيان تقوم على مسافات حقيقية يتمُّ قياسها، ولكن هذا لا يلزم أن يكون كذلك، كما هو الحال عادةً في تطوير نظم المعلومات. كما يمكن اســتخدامُ المتوســطات المرجّحة لتحديد هذه المســافات. فمثلًا في مشروع تطوير IS، قد تكون الوحدات الفردية للنظام مرتبطة بالتشــابه بين مدخلاتها، والمخرجات، والعمليات، والبيانات المُحدَّدة المُستَخدَمة. ثم يتمُّ تجميع هذه العوامل، متزاوجةً حسب العنصر، في مقياس مسافة واحدة.

الخوارزمية التجميعية K-means: يمكن القول بأن خوارزمية K-means (حيث تشــير k إلى عددٍ مُحدَّد مسبقاً من المجموعات) هو أكثر الخوارزميات التجميعية التي يُشار إليها. كما أن لها جذورها في التحليل الإحصائي التقليدي. وكما يوحي الاسم؛ فإن هذه الخوارزمية تقوم بتعيين كل نقطة بيانات (العميل، الحدث، الكائن... إلخ) إلى المجموعة التي يكون مركزها (التي يُطلق عليها أيضًا النقطة الوســطى) هو الأقرب. ويتم حســابُ المركز كمتوســط جميع النقاط في المجموعة؛ أي إنَّ إحداثياتها هي المتوســط الحسابي لكلِّ بُعد بشــكلٍ منفصلٍ على كل النقاط في المجموعة. ويوضِّح الشكل ٤-١٣ خطوات هذا الخوارزمية والتي سنذكرها فيما يلي:

- خطوة التهيئة: اختر عدد المجموعات (أيْ: قيمة K).
- الخطوة ١: إنشاء نقاط عشوائية k عشوائيًّا كمراكز مجموعة أولية.
- الخطوة ٢: قم بتعيين كل نقطة إلى مركز المجموعة الأقرب.
- الخطوة ٣: إعادة إنشاء مراكز المجموعات الجديدة.
- خطــوة التكــرار: كرر الخطوتين ٢ و٣ حتى يتم اســتيفاء معيار تقارب (عــادةً ما يصبح تعيين النقاط إلى المجموعات ثابتًا).



شكل ٤-١٣: توضيح بياني للخطوات في خوارزمية K-Means

التنقيب في قاعدة الترابط:

يُعَدُّ التنقيب في قاعدة الترابط (المعروف أيضًا باســم تحليل التقارب أو تحليل ســلة الســوق) أســلوبًا شــائعًا للتنقيب في البيانات؛ إذ يُســتَخدَم عادةً كمثالٍ لتوضيح ما الــذي يعنيه التنقيب في البيانات، وما يمكن أن يفعله لجمهور أقل خبرة من الناحية التقنية. وقد يكون معظمكم قد ســمع عن العلاقة الشهيرة (أو سوء السمعة؛ فهذا يعتمد على نظرتكم إليها) التي تفصل بين مبيعات البيرة والحفاضات في محلات البقالة. وكما تقول القصة؛ فإن سلسلة متاجر كبيرة (ربما Walmart، وربما لا؛ حيث لا يُوجَد توافق في الآراء بشأن سلسلة السوق الفائقة التي كانت عليها) قامت بتحليل عادات الــشراء لدى العمــلاء، ووجدت علاقة ارتباط ذات دلالة إحصائية بين مشــتريات البيرة ومشــتريات الحفاضات. وقد وُجد أن السبب في ذلك هو أن الآباء (الذين يفترض أنهم شباب) كانوا يتوقفون في الســوبر ماركت لشراء حفاضات لأطفالهم (خاصةً أيام الخميس)، ولأنهم لم يَعُد بإمكانهم النزول إلى ســاحة الرياضة في كثيرٍ من الأحيان، مثلما لم يعد بإمكانهم شراء البيرة أيضًا. ونتيجةً لذلك، يُزعم أن سلسلة المتاجر الكبرى وضعت الحفاضات بجانب البيرة؛ مما أدى إلى زيادة مبيعات كليهما.

ويهدف استخراج قاعدة المجموعات في جوهره إلى إيجاد علاقات مثيرة للاهتمام (الصلات) بين المتغيرات (البنود) في قواعد البيانات الكبيرة. وبسبب تطبيقه الناجح لمشكلات تجارة التجزئة؛ فإنه يُطلَق عليه عادةً تحليل سلة السوق. وتتمثل الفكرة الرئيسية في تحليل سلة السوق في تحديد العلاقات القوية بين المنتجات المختلفة (أو الخدمات) التي يتمُّ شراؤها عادةً معًا (تظهر في نفس السلة معًا، إما سلة مادية في متجر للبقالة أو سلة افتراضية على موقع للتجارة الإلكترونية). فمثلًا يشتري ٦٥٪ من أولئك الذين يشترون تأمينًا شاملًا للسيارات التأمين الصحي أيضًا. و٨٠٪ من أولئك الذين يشترون الكتب عبر الإنترنت يشترون الموسيقى عبر الإنترنت أيضًا؛ بالإضافة إلى أن ٦٠٪ من أولئك الذين لديهم ارتفاعٌ في ضغط الدم وزيادة الوزن؛ لديهم ارتفاع كوليسترول في الدم؛ و٧٠٪ من العملاء الذين يشترون حاسباً محمولاً وبرامج الحماية من الفيروسات يشترون أيضًا خطط خدمة مُوسَّعة.

وتتمثل المدخلات في تحليل سلة السوق في بيانات معاملات نقطة البيع البسيطة؛ إذ يتمُّ تبويب عددٍ من المنتجات أو الخدمات التي تمَّ شراؤها معًا (تمامًا مثل محتوى إيصال الشراء) في حالة معاملة واحدة. وتُعَد نتيجة التحليل بمثابة معلومات لا تُقدَّر بثمن والتي يمكن استخدامها لفهم سلوك الشراء لدى العملاء بشكلٍ أفضل لتحقيق أقصى قدرٍ من الرِّبح من المعاملات التجارية؛ إذ تستفيد الأعمال التجارية من هذه المعرفة عبر:

١- وَضْع العناصر بعضها بجانب بعض؛ لجعلها أكثر ملاءمةً للعملاء لالتقاطها معًا وعدم نسيان شراء واحدة عند شراء الأخرى (زيادة حجم المبيعات).

٢- الترويج للبنود كحزمة (لا تضع واحدةً للبيع إذا كان الآخرون معروضين للبيع).

٣- وضعها بعضها بعيدًا عن بعض؛ بحيث يكون على العميل المشي في الممرات للبحث عنه؛ وذلك من خلال القيام بذلك من المحتمل رؤية وشراء سلع أخرى.

وتشتمل تطبيقات تحليل سلة السوق على التسويق المتقاطع، والبيع المتقاطع، وتصميم المتجر، وتصميم الكتالوج، وتصميم موقع التجارة الإلكترونية؛ بالإضافة إلى تحسين الإعلان عبر الإنترنت، وتسعير المنتج؛ فضلًا عن تكوين المبيعات/ الترويج؛ إذ يساعدُ تحليل سلة السوق في جوهره الشركات على استنتاج احتياجات العملاء وتفضيلاتهم من أنماط الشراء. وخارج نطاق عالم الأعمال؛ فإن قواعد المجموعات تُستَخدَم بنجاح لاكتشاف العلاقات بين الأعراض والأمراض والتشخيص وخصائص وعلاجات المريض (التي يمكن استخدامها في نظام دَعْم القرار الطبي) والجينات ووظائفها (التي يُمكِن استخدامها في مشاريع الجينوميات)، من بين أمورٍ أخرى. وفيما يلي بعض المناطق والاستخدامات الشائعة لاستخراج قاعدة المجموعات:

- معاملات البيع: يُمكِن استخدامُ مجموعات من منتجات التجزئة التي تم شراؤها معًا لتحسين موضع المنتج في الطابق المخصَّص للمبيعات (وضع المنتجات التي تتماشى بعضها مع بعض) والتسعير الترويجي للمنتجات (وليس وجود عروض ترويجية على كلٍّ من المنتجات التي يتمُّ شراؤها في الغالب معًا).
- معاملات بطاقات الائتمان: توفر العناصر المشتراة من خلال بطاقة ائتمانية بيانات عن منتجات أخرى يُرجَّح قيامُ العميل بشرائها أو استخدام أرقام بطاقات الائتمان.
- الخدمات المصرفية: يُمكِن استخدام أنماط الخدمات المتسلسلة التي يستخدمها العملاء (فحص الحساب متبوعًا بحساب التوفير) لتحديد الخدمات الأخرى التي قد يكونون مهتمين بها (حساب استثماري).
- منتجات خدمات التأمين: يمكن استخدامُ حزم منتجات التأمين التي يشتريها العملاء (التأمين على السيارات يتبعها التأمين المنزلي) لاقتراح منتجات تأمين إضافية (التأمين على الحياة)، أو مزيج غير عادي من مطالبات التأمين يمكن أن يكون علامةً على الاحتيال.
- خدمات الاتصالات: تساعد مجموعات الخيارات التي تمَّ شراؤها بشكلٍ عام (مثل: انتظار المكالمات، وهوية المتصل، والاتصال الثلاثي) على تكوين حزم المنتجات بشكلٍ أفضل لزيادة الإيرادات. وينطبق الشيء نفسه على مُزوِّدي خدمات الاتصالات المتعددة القنوات من خلال الهاتف والتلفزيون وعروض خدمة الإنترنت.
- السجلات الطبية: يمكن أن تشير توليفات معينة من الحالات إلى زيادة خطر حدوث مضاعفات مختلفة؛ أو يمكن ربط بعض إجراءات العلاج في بعض المرافق الطبية بأنواع معينة من العدوى.

إن أفضل سؤالٍ ينبغي طرحه فيما يخصُّ الأنماط/ العلاقات التي يمكن أن تكتشفها قواعد مجموعات الاستخراج هو «هل كل قواعد الترابط مثيرة للاهتمام ومفيدة؟» وللإجابة عن سؤال من هذا القبيل؛ فإن قواعد ترابط الاستخراج تستخدم مقياسين شائعين: الدعم والاعتدال والرفع. وقبل تعريف هذه المصطلحات؛ دعنا نحصل على القليل من التقنية من خلال إظهار شكل قاعدة الارتباط:

{حاسب محمول، برامج مكافحة الفيروسات} ١ {خطة الخدمة الموسعة} [٣٠٪، ٧٠٪].

حيث ترتبط X (المنتجات أو الخدمة، تسمى الجانب الأيسر، LHS، أو سابقة) بـ Y (المنتجات أو الخدمة، وتسمى الجانب الأيمن، RHS، أو مرافقه). وS هو الدعم، وC هي الثقة لهذه القاعدة الخاصة. فيما يلي الصيغ البسيطة لـ Supp وConf وLift.

$$Support = Supp(X \Rightarrow Y) = \frac{\text{number of baskets that contains both X and Y}}{\text{total number of baskets}}$$

$$Confidence = Conf(X \Rightarrow Y) = \frac{Supp(X \Rightarrow Y)}{Supp(X)}$$

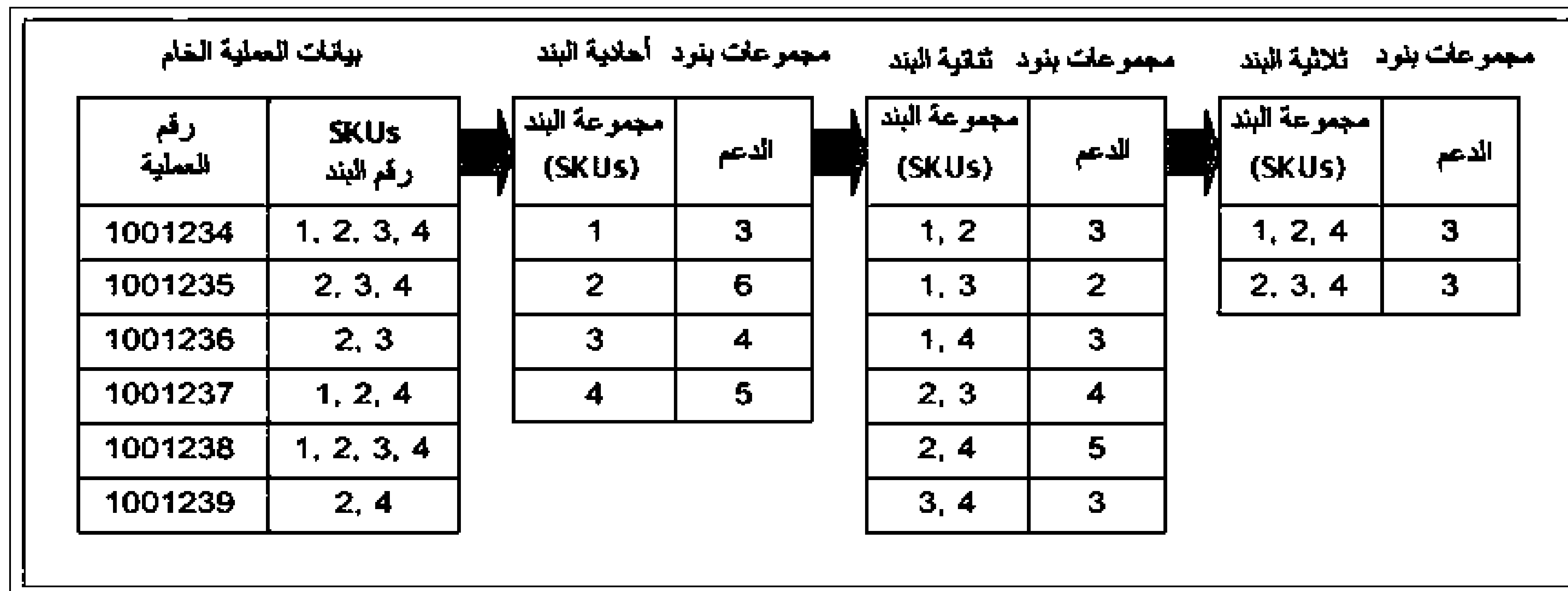
$$Lift(X \Rightarrow Y) = \frac{Conf(X \Rightarrow Y)}{Expected\ Conf(X \Rightarrow Y)} = \frac{\frac{S(X \Rightarrow Y)}{S(X)}}{\frac{S(X) \cdot S(Y)}{S(X)}} = \frac{S(X \Rightarrow Y)}{S(X) \cdot S(Y)}$$

إن الدَّعم (S) لمجموعة من المنتجات هو مقياس مدى تكرار ظهور هذه المنتجات أو الخدمات (مثل LHS + RHS = الحاسب المحمول وبرامج مكافحة الفيروسات وخطة الخدمة الموسَّعة) في نفس المعاملة؛ أيْ: نسبة المعاملات في مجموعة البيانات التي تحتوي على جميع المنتجات أو الخدمات المذكورة في قاعدة مُحدَّدة. وفي هذا المثال كانت ٣٠٪ من جميع المعاملات في قاعدة بيانات المتجر الافتراضية تحتوي على جميع المنتجات الثلاثة الموجودة في بطاقة مبيعات واحدة. وثقة القاعدة هي مقياس مدى تلازم المنتجات أو الخدمات في RHS (الناتجة) مع المنتجات أو الخدمات على نظام LHS (سابقاً)، أيْ: نسبة المعاملات التي تتضمَّن نظام LHS في حين تتضمَّن أيضًا RHS. وبعبارة أخرى؛ فإن الاحتمال المشروط هو العثور على RHS للقاعدة الموجودة في المعاملات التي يُوجَد بها بالفعل نظام LHS للقاعدة. وقيمة الرفع لقاعدة الارتباط هي نسبة ثقة القاعدة والثقة المتوقعة من القاعدة. ويتمُّ تعريف الثقة المتوقعة للقاعدة على أنها نتاج قيم دعم LHS وRHS مقسومًا على دعم نظام LHS.

وتتوفر عِدَّة خوارزميات لاكتشاف قواعد الترابط؛ إذ تتضمَّن بعض الخوارزميات المعروفة Apriori وEclat بالإضافة إلى FP-Growth. وتقوم هذه الخوارزميات فقط بنصف المهمة، وهو تحديد العناصر المتكررة في قاعدة البيانات. وحالما يتمُّ تحديد العناصر المتكررة؛ فإنها تحتاج إلى تحويلها إلى قواعد مع الأجزاء السابقة واللاحقة؛ إذ إنَّ تحديد القواعد من المواد المتداولة هو عملية مطابقة مباشرة؛ لكن العملية قد تستغرق وقتًا طويلًا مع قواعد بيانات المعاملات الكبيرة. وعلى الرغم من أنه يمكن أن يكون هناك العديد من العناصر في كلِّ مقطع من القاعدة؛ فإنَّ الجزء الذي يترتب على ذلك عادةً ما يحتوي على عنصر واحد. وفي القسم التالي سوف نتناول شرح إحدى الخوارزميات الأكثر شيوعًا لتحديد العناصر المتكررة.

خوارزمية Apriori: وهو النظام الأكثر استخدامًا لاكتشاف قواعد الارتباط. وبالنظر إلى مجموعة من العناصر (مثل مجموعات من معاملات البيع بالتجزئة، وكل عنصر من العناصر الفردية التي تمَّ شراؤها)، تحاول هذه الخوارزمية العثور على مجموعاتٍ فرعية شائعة على الأقل كحَدٍّ أدنى لعدد العناصر (أيْ: إنها تتوافق مع الحد الأدنى من الدعم). ويستخدم Apriori أسلوبًا من القاعدة إلى القمة؛ إذ يتمُّ توسيعُ مجموعات فرعية متكررة عنصرًا واحدًا في كلِّ مرة (وهي طريقة تُعرَف باسم جيل المرشح؛ إذ يزدادُ حجم المجموعات الفرعية المتكررة من مجموعات فرعية ذات عنصر واحد إلى مجموعاتٍ فرعية من فئتين، ثم مجموعات فرعية ثلاثية العناصر... الخ)، ويتمُّ اختبار مجموعاتٍ من المرشحين في كلِّ مستوى ضد البيانات؛ للحصول على الحَدِّ الأدنى من الدعم. وتنتهي الخوارزمية عند عدم العثور على ملحقات إضافية ناجحة.

وكمثالٍ توضيحي: ضَعْ في اعتبارك ما يلي: يتتبع متجر البقالة عمليات نقل المبيعات بواسطة SKU (وحدة حفظ المخزون)، وبالتالي يعرف العناصر التي يتمُّ شراؤها عادةً معًا. يظهر في الشكل ٤-١٤ قاعدة بيانات المعاملات، إلى جانب الخطوات اللاحقة لتحديد العناصر المتكررة. وكلُّ رمز SKU في قاعدة بيانات المعاملة يتحول إلى منتج، مثل: «butter = 1»، و«bread = 2»، و«water = 3»، وما إلى ذلك. تتمثل الخطوة الأولى في Apriori في حساب الترددات (أيْ: الدعامات) لكلِّ عنصر (عناصر من عنصرٍ واحد). ولهذا المثال المبسَّط بشكلٍ كبير، دعنا نعيِّن الحد الأدنى للدعم إلى ٣ (أو ٥٠٪؛ مما يعني أن العنصر يُعَدُّ عنصرًا متكررًا إذا تمَّ عرضه في ٣ من أصل ٦ معاملات على الأقل في قاعدة البيانات). ونظرًا لأن جميع العناصر المكونة من عنصر واحد تحتوي على ٣ على الأقل في عمود الدعم؛ فكلها تُعَدُّ عناصر متكررة. ومع ذلك؛ فإنه إذا لم تكن أيٌّ من عناصر المجموعة الواحدة متكررة؛ فلن يتمَّ إدراجها كعضو محتمل في زوجين من النقاط محتملين. وبهذه الطريقة، يُحرِّر Apriori شجرة جميع العناصر الممكنة. وكما يوضح الشكل ٤-١٤؛ فإنه باستخدام عناصر مجموعة واحدة، يتمُّ إنشاء جميع عناصر العناصر الممكنة، ويتمُّ استخدام قاعدة بيانات المعاملة لحساب قيم الدعم الخاصة بها. ونظرًا لأن عنصر العناصر المكون من عنصرين {١، ٣} يحتوي على دَعْم أقل من ٣؛ فلا يجب تضمينه في العناصر المتكررة التي سيتمُّ استخدامها لإنشاء مجموعات العناصر التالية (عناصر من العناصر الثلاثة). وتبدو هذه الخوارزمية بسيطةً بشكلٍ خادع؛ ولكن فقط لمجموعات البيانات الصغيرة. أما في مجموعات البيانات الأكبر بكثير - خاصة تلك التي تحتوي على كميات هائلة من العناصر الموجودة بكميات منخفضة وكميات صغيرة من العناصر الموجودة في كميات كبيرة - فإن البحث والحساب يصبح عملية مكثفة من حيث الحساب.



بيانات العملية الخام

رقم العملية	SKUs رقم البند
1001234	1, 2, 3, 4
1001235	2, 3, 4
1001236	2, 3
1001237	1, 2, 4
1001238	1, 2, 3, 4
1001239	2, 4

مجموعات بنود أحادية البند

مجموعة البند (SKUs)	الدعم
1	3
2	6
3	4
4	5

مجموعات بنود ثنائية البند

مجموعة البند (SKUs)	الدعم
1, 2	3
1, 3	2
1, 4	3
2, 3	4
2, 4	5
3, 4	3

مجموعات بنود ثلاثية البند

مجموعة البند (SKUs)	الدعم
1, 2, 4	3
2, 3, 4	3

شكل ٤-١٤: تحديد مجموعات البنود المتكررة في خوارزمية Apriori

أسئلة مراجعة على القسم ٤-٥:

١- حَدِّد على الأقل ثلاث طرقٍ من الطرق الرئيسة للتنقيب في البيانات.

٢- اذكــر أمثلــةً عن مواقف يمكن أن يكون فيها التصنيف أســلوباً مناســباً للتنقيب في البيانات. واذكر أمثلةً على الحالات التي يكون فيها الانحدارُ أسلوبًا مناسبًا للتنقيب في البيانات.

٣- اذكر على الأقل اثنين من تقنيات التصنيف مع تعريفٍ موجزٍ لهما.

٤- ما هي بعض المعايير لمقارنة واختيار أفضل تقنية للتصنيف؟

٥- صِف باختصار الخوارزمية العامة المستخدمة في أشجار القرار.

٦- عرِّف مؤشر جيني. واذكر ماذا يقيس؟

٧- ما هو نموذج المجموعة في مجال التنقيب في البيانات؟ ما هي مزايا وعيوب نماذج المجموعة؟

٨- اذكر أمثلةً للمواقف التي سيكون فيها التحليلُ التجميعي تقنيةً مناسبةً لاستخراج البيانات.

٩- ما هو الفرق الرئيس بين التحليل التجميعي والتصنيف؟

١٠- اذكر بعض طرق التحليل التجميعي؟

١١- اذكر أمثلةً على الحالات التي يكون فيها الارتباط تقنية مناسبة للتنقيب في البيانات.

٤-٦ أدوات برمجيات التنقيب في البيانات:

يوفر العديدُ من مورِّدي البرامج أدواتٍ قويةً للتنقيب في البيانات. من أمثلة هؤلاء الموردين:

- شركة IBM SPSS Modeler) MBI، والمعروفة سابقًا باسم SPSS PASW Modeler وClementine).

- شركة SAS (رائدة المشروعات).
- شركة Dell (Statistica، والمعروفة سابقًا باسم StatSoft Statistica Data Miner).
- شركة PAS (Infinite Insight، والمعروف سابقًا باسم KXEN Infinite Insight).
- أنظمة سالفورد (CART, MARS, TreeNet, RandomForest).
- شركة Angoss (KnowledgeSTUDIO, KnowledgeSEeker).
- شركة Megaputer (PolyAnalyst).

بشكل ملحوظ ولكن ليس من المستغرب، يتم تطوير أدوات التنقيب في البيانات الأكثر شيوعًا واستخدامًا من قبل شركات البرمجيات الإحصائية الراسخة إلى حد كبير (مثل SAS, SPSS, StatSoft)؛ لأن الإحصاءات هي أساس التنقيب في البيانات، وهذه الشركات لديها وسائل فعالة من حيث التكلفة تقوم بتطويرها في أنظمة الإستخراج الشامل للبيانات. كما أن معظم مُورِّدي أدوات ذكاء الأعمال (مثل: IBM Cognos، وOracle Hyperion، وSAP Business Objects، وTableau، وTibco، وQlik، وMicroStrategy، وTeradata، وMicrosoft) لديهم أيضًا مستوى معين من قدرات التنقيب في البيانات المُدمَجة في عروض البرامج الخاصة بهم. ولا تزال أدوات ذكاء الأعمال هذه تركِّز في المقام الأول على النمذجة المتعددة الأبعاد وتصوير البيانات، ولا تُعَدُّ منافسةً مباشرة لمورِّدي أدوات التنقيب في البيانات.

إضافةً إلى هذه الأدوات التجارية؛ تتوفر العديدُ من أدوات برمجيات جمع البيانات المفتوحة المصدر أو المجانية على الإنترنت. وبشكلٍ بديهي - ولا سيما في الإدارات التربوية - تُعَدُّ Weka أداة التنقيب في البيانات المجانية والمفتوحة الأكثر انتشارًا، والتي تمَّ تطويرها من قِبل عددٍ من الباحثين من جامعة Waikato في نيوزيلندا (يمكن تنزيل الأداة من cs.waikato.ac.nz/ml/weka). تحتوي Weka على عددٍ كبيرٍ من الخوارزميات لمهامّ مختلفة للتنقيب في البيانات ولديها واجهة مُستخدِم سهلة الاستخدام. وقد ظهر في الآونة الأخيرة عددٌ من الأدوات المجانية المفتوحة المصدر وذات قدرة عالية على التنقيب في البيانات: يقود المجموعة KNIME (knime.org) وRapidMiner (quickminer.com). فلديهما واجهات مُستخدِم بيانية مُحسَّنة، وأيضًا بهما عددٌ كبيرٌ من الخوارزميات، ومجموعة متنوعة من مَيْزات التمثيل البصري للبيانات، كل تلك الخصائص تميزهما عن بقية الأدوات المجانية. ويُعَدُّ هذان البرنامجان المجانيان أيضًا برنامجين أساسيين (أيْ: يُمكِن تشغيلهما على نظامي تشغيل Windows وMac). ومع التغيير الأخير في عروضها؛ قامت RapidMiner بإنشاء نسخة مُصغَّرة من أداة التحليلات الخاصة بهم مجانًا

(أيْ: نسـخة مجتمعية)؛ وذلـك في أثناء تصنيع المنتـج التجاري الكامل. لذلـك؛ فبمجرد إدراج RapidMiner ضمن فئة الأدوات المجانية/ مفتوحة المصدر، غالبًا ما يتم إدراجها في قائمة الأدوات التجاريـة. يُعَدُّ الاختلافُ الرئيس بين كلٍّ من الأدوات التجارية (مثل: SAS Enterprise Miner، وIBM SPSS Modeler، وStatistica) والأدوات المجانيـة (مثـل: Weka، وRapidMiner، وKNIME) هو الكفاءة الحسابية. وقد تستغرق مهمة التنقيب في البيانات نفسها التي تتضمن مجموعـة بيانات كبيرة وغنية بالميزات وقتاً أطول بكثـير لتكتمل مع أدوات البرمجيات المجانية، وبالنسـبة لبعض الخوارزميات، قد لا تكتمل المهمة (أيْ: تتعطل بسـبب الاسـتخدام غير الفعَّال لذاكرة الحاسب). يعرض الجدول ٤-٢ بعضًا من المنتجات الرئيسة والمواقع الخاصة بها.

هناك مجموعةٌ من قدرات ذكاء الأعمال والتحليلات التي أصبحت أكثر شـيوعًا بشكلٍ متزايد لدراسـات التنقيـب في البيانات، وهـي Microsoft SQL Server (فقد تـمَّ تضمينها لإمكانات تحليلات أكثر بشكلٍ متزايد، مثل ذكاء الأعمال ووحدات النمذجة التنبؤية؛ وذلك بدءًا من إصدار SQL Server 2012)؛ إذ يتمُّ عمل مسـتودعات البيانات والنماذج في نفس بيئة قاعدة البيانات العلائقية؛ مما يجعل إدارة النماذج مهمةً أسـهل بكثير. يعمل تحالف Microsoft Enterprise كمصدر عالمـي للوصـول إلى مجموعة برامـج Microsoft SQL Server للأغـراض الأكاديمية، كالتدريـس والأبحاث. وقد تمَّ إنشـاءُ هـذا التحالف؛ لتمكين الجامعـات في جميع أنحاء العالم من الوصول إلى تقنية المؤسسـات دون الحاجة إلى صيانة الأجهـزة والبرامج الضرورية في مبانيها الخاصـة. كما يوفر التحالف مجموعةً واسـعةً مـن أدوات تطوير ذكاء الأعـمال (كالتنقيب في البيانـات، بناء المكعبات، إعداد تقارير الأعمال) إضافـةً إلى عددٍ من مجموعات البيانات الكبيرة والواقعية من Sam's Club وDillard's وTyson Foods. يُعَدُّ تحالف Microsoft Enterprise مجانيًّا ولا يمكن اسـتخدامه إلا للأغراض الأكاديمية. تسـتضيف كلية Sam M. Walton للأعمال بجامعة Arkansas النظام المؤسـسي وتسـمح لأعضاء التحالف وطلابه بالوصول إلى هذه الموارد باسـتخدام اتصالٍ بسيط عن بعد بسطح المكتب. ويمكن معرفة تفاصيل الانضمام لهذا التحالف وكذلك الدروس سهلة المتابعة وبعض الأمثلة في walton.uark.edu /enterprise.

جدول ٤-٢: برامج مختارة للتنقيب في البيانات

اسم المنتج	الموقع (URL)
IBM SPSS Modeler	www-01.ibm.com/software/analytics/spss/products/modeler/
IBM Watson Analytics	ibm.com/analytics/watson-analytics/
SAS Enterprise Miner	sas.com/en_id/software/analytics/enterprise-miner.html
Dell Statistica	statsoft.com/products/statistica/product-index
PolyAnalyst	megaputer.com/site/polyanalyst.php
CART, RandomForest	salford-systems.com
Insightful Miner	solutionmetrics.com.au/products/iminer/default.html
XLMiner	solver.com/xlminer-data-mining
SAP InfiniteInsight (KXEN)	help.sap.com/ii
GhostMiner	fqs.pl/ghostminer
SQL Server Data Mining	msdn.microsoft.com/en-us/library/bb510516.aspx
Knowledge Miner	knowledgeminer.com
Teradata Warehouse Miner	teradata.com/products-and-services/Teradata-warehouse-miner/
Oracle Data Mining (ODM)	oracle.com/technetwork/database/options/odm/
FICO Decision Management	fico.com/en/analytics/decision-management-suite/
Orange Data Mining Tool	orange.biolab.si/
Zementis Predictive Analytics	zementis.com

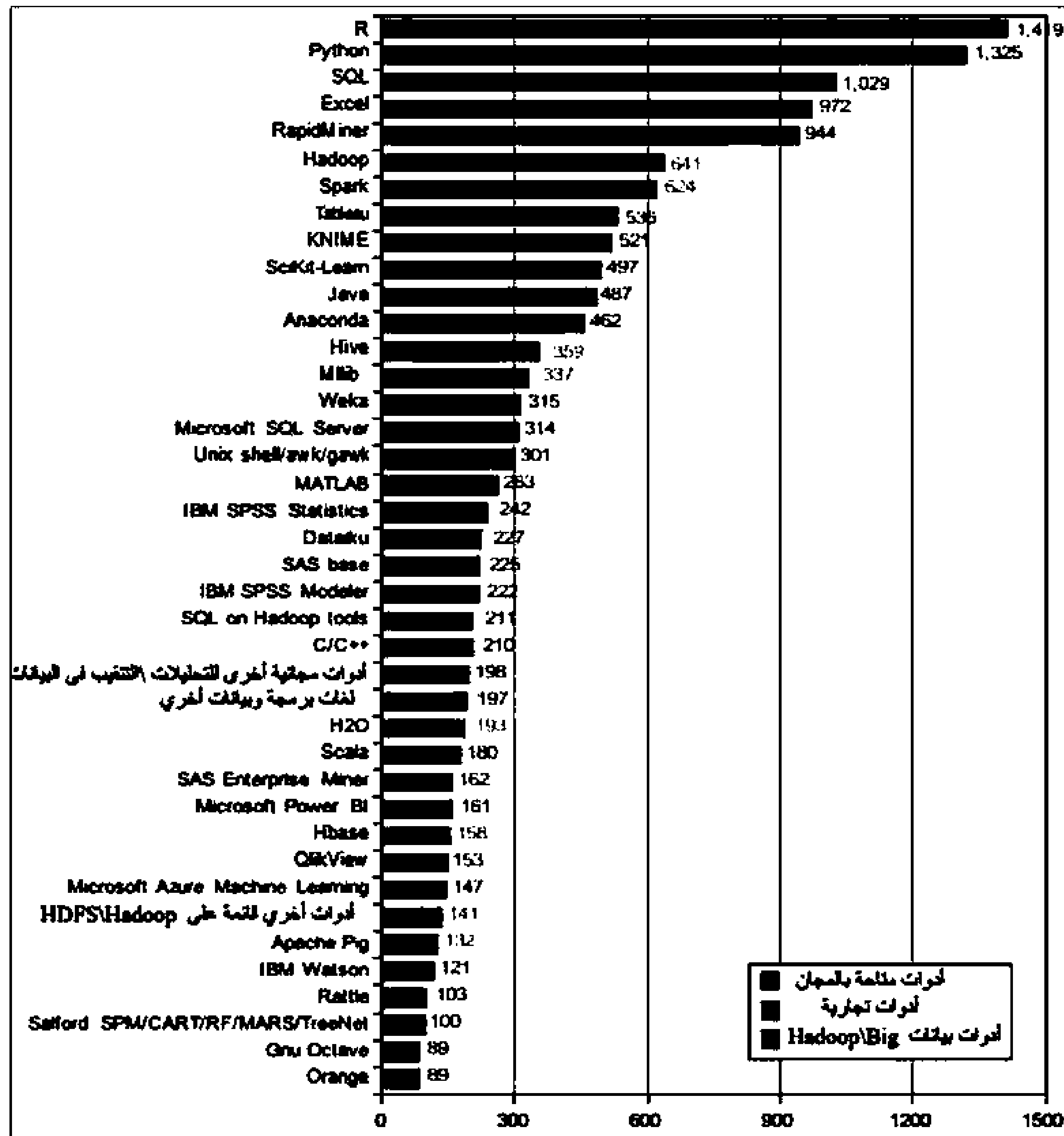
في مايو ٢٠١٦م، أجرت KDnuggets.com اســتطلاع البرامج الســنوي الثالث عشر بشأن السؤال التالي: «ما هي البرامج التي استخدمتها في التحليلات، والتنقيب في البيانات، وعلم البيانات، ومشاريع تعلُّم الآلة خلال الـ ١٢ شــهرًا الماضية؟». وقد حظي الاستطلاع بمشاركةٍ رائعة من مجتمع التحليلات وعلم البيانات والمورِّدين؛ مما جذب ٢٨٨٩ ناخبًا اختاروا عددًا قياسيًّا بلغ ١٠٢ من الأدوات المختلفة.

وقد أثمر هذا الاستطلاع عن بعض النتائج المثيرة للاهتمام منها:

- لا تزال R هي الأداة الرئيسة، بحصة تبلغ ٤٩٪ من الأسهم (بعد أن كانت ٤٦,٩٪ في عام ٢٠١٥م)، ولكــن اســتخدام Python ازداد بسرعة واكتســب ما يقرب من R بحصة ٤٥,٨٪ من الأســهم (والتي توقفت عند ٣٠,٣٪ في ٢٠١٥م).
- لا تزال RapidMiner هي المنصة العامة الأكثر انتشارًا بشأن التنقيب في البيانات/ علم البيانات، بنسبة بلغت ٣٣٪ من الأسهم. ومن الأدوات البارزة التي تحظى بأكبر قدرٍ من النمو في الانتشار هــي Dato، وDataiku، وMLlib، وH2O، وAmazon Machine Learning، و-scikit learn، وIBM Watson.
- ينعكس الخيار المتزايد للأدوات في الاســتخدام الأوسع؛ إذ بلغ متوسط عدد الأدوات المستخدمة ٦,٠ (مقابل ٤,٨ في مايو ٢٠١٥م).
- ارتفــع اســتخدام أدوات Hadoop/ البيانات الكبــيرة إلى ٣٩٪، بعد أن كان ٢٩٪ في عام ٢٠١٥م و١٧٪ في عام ٢٠١٤م، مدفوعًا من قبل Apache Spark وMLlib (مكتبة Spark لتعلم الآلة) وH2O. والتي أدرجناها ضِمن أدوات البيانات الكبيرة.
- كانت المشاركة حسب المنطقة؛ وذلك كالتالي: الولايات المتحدة / كندا (٤٠٪)، أوروبا (٣٩٪)، آسيا (٩,٤٪)، أمريكا اللاتينية (٥,٨٪)، أفريقيا / الشرق الأوسط (٢,٩٪)، أستراليا / نيوزيلندا (٢,٢٪).
- في هذا العام، اســتخدم ٨٦٪ من المصوتين برامج تجارية، في حين استخدم ٧٥٪ البرامج المجانية. واســتخدم حوالي ٢٥٪ فقط البرامج التجارية، و١٣٪ فقط استخدموا البرامج المفتوحة المصدر / المجانية. واســتخدم الأغلبية بمقدار ٦١٪ كلًا من البرامج المجانية والتجارية على حَدٍّ ســواء، في حين كانت ٦٤٪ في عام ٢٠١٥م.
- بالنســبة للســنة الثانية، شمل اســتطلاع KDnuggets.com أدوات التعلُّم العميق. ففي هذا العام، اســتخدم ١٨٪ من المصوِّتين أدوات التعلم العميق، أيْ: ضِعف نسبة عام ٢٠١٥م، والتي كانــت ٩٪، وبهذا قفز موقــع Google Tensorflow إلى المركز الأول، متخطيًــا النظام البيئي Theano/Pylearn2 صاحب الصدارة في العام الماضي.

- في مجتمـع لغات البرمجة، ارتفعت شـعبية كلٍّ مـن Python وJava وUnix وScala؛ في حين انخفضت شعبية كلٍّ من C/C++، وPerl، وJulia، و# F، وClojure، وLisp.

وللحَـدِّ مـن التحيُّز في التصويت المتعَـدِّد؛ قام موقع KDnuggets.com في هذا الاسـتطلاع بالتحقـق من البريـد الإلكتروني؛ وذلك بهدف جَعْل النتائج أكثر تمثيـلًا للواقع في عالم التحليلات. ويُظهِر في الشكل ٤-١٥ نتائج لأفضل ٤٠ من الأدوات البرمجية (وذلك وفقًا لإجمالي عدد الأصوات المسـتلمة). كما يقوم الرسـم البياني الأفقي أيضًا بالتمييز بين الأدوات المجانية/ مفتوحة المصدر والتجارية وبيانات Hadoop / البيانات الضخمة باستخدام مخطط الترميز ذي الألوان.



Source: used with permission from KDnuggets

شكل ٤-١٥: الأدوات البرمجية الشائعة للتنقيب في البيانات

حالة عملية ٤-٦

التنقيب في البيانات ينتقل إلى هوليوود: توقع النجاح المالي للأفلام

تتعلّق الحالة العملية ٤-٦ بدراسةٍ بحثية؛ إذ يتمُّ استخدام عددٍ من الأدوات البرمجية وتقنيات جمع البيانات لبناء نماذج التنقيب في البيانات؛ للتنبؤ بالنجاح المالي (إيرادات شباك التذاكر) لأفلام هوليوود في حين أنها ليست أكثر من مجرد أفكار.

يُعَدُّ توقع إيرادات شباك التذاكر (أيْ: النجاح المالي) لصورةٍ متحركةٍ معينةٍ؛ مشكلةً مثيرةً للاهتمام ومليئةً بالتحديات. ووفقاً لبعض خبراء المجال؛ فإن صناعة السينما تُعَدُّ «أرض العقبات والمخاطر الجامحة»؛ بسبب الصعوبة المرتبطة بالتنبؤ بالطلب على المنتج؛ مما يجعل أعمال السينما في هوليوود محفوفةً بالمخاطر. ودعمًا لمثل هذه الملاحظات، يقول Jack Valenti (الرئيس والمدير التنفيذي السابق لاتحاد الصور المتحركة الأمريكية): «لا أحد يستطيع أن يخبرك كيف سينتج الفيلم في السوق، ليس قبل أن يفتح الفيلم في مسرح مظلم، والشرر يطير بين الشاشة والجمهور». وقد كانت الصحف والمجلات التجارية الخاصة بصناعة الترفيه مليئةً بالأمثلة والتصريحات والخبرات التي تدعم مثل تلك المطالب.

وكما هو الحال بالنسبة للعديدِ من الباحثين الذين حاولوا إلقاء الضوء على هذه المشكلة المليئة بالتحَدِّيات في العالم الحقيقي، كان كلٌّ من Ramesh Sharda، وDursun Delen يستكشفان استخدام التنقيب في البيانات؛ للتنبؤ بالأداء المالي للصورة المتحركة في شباك التذاكر قبل أن يدخل الإنتاج (في حين أن الفيلم ليس أكثر من فكرة مفاهيمية). وفي نماذج التنبؤ ذات الانتشار الكبير، يتمُّ تحويل مشكلة التنبؤ (أو الانحدار) إلى مشكلة تصنيف؛ بمعنى أنه بدلًا من التنبؤ بتقدير نقطة لإيرادات شباك التذاكر؛ يتمُّ تصنيف أيِّ فيلم وفقًا لإيراداته ووضعه في إحدى الفئات التسع، والتي تتراوح من الهبوط الشديد «flop» إلى الإقبال «blockbuster»؛ مما يجعل المشكلة مشكلة تصنيف متعددة الحدود. ويُوضِّح الجدول ٤-٣ تعريف الفئات التسع من حيث نطاق إيصالات شباك التذاكر.

TABLE 4.3 Movie Classification Based on Receipts

Class No.	1	2	3	4	5	6	7	8	9
Range (in millions of dollars)	> 1 (Flop)	> 1 < 6 10	> 10 < 20	> 20 < 6 40	> 40 < 6 65	> 65 < 6 100	> 100 < 6 150	> 150 < 6 200	> 200 (Blockbuster)

البيانات:

تمَّ جمع البيانات من مجموعةٍ متنوعةٍ من قواعد البيانات المتعلقة بالأفلام (مثل: Showbiz، وIMDb، وIMSDb، وAllMovie، وBoxofficeMojo، وغيرهــا) وقد تمَّ دمجها جميعًا في مجموعة بيانــات واحدة؛ إذ تحتوي مجموعة البيانات الخاصة بأكثر النماذج تطورًا في الآونة الأخيرة على ٢٦٣٢ فيلماً تمَّ إصدارها بين عامي ١٩٩٨ و٢٠٠٦. ويعرض الجدول ٤-٤ ملخصاً للمتغيرات المســتقلة إلى جانــب مواصفاتها. ولمزيدٍ من التفاصيل الوصفية ومبرِّرات تضمين هذه المتغيرات المســتقلة؛ يتمُّ الاطلاع على كلٍّ من Sharda وDelen (2006).

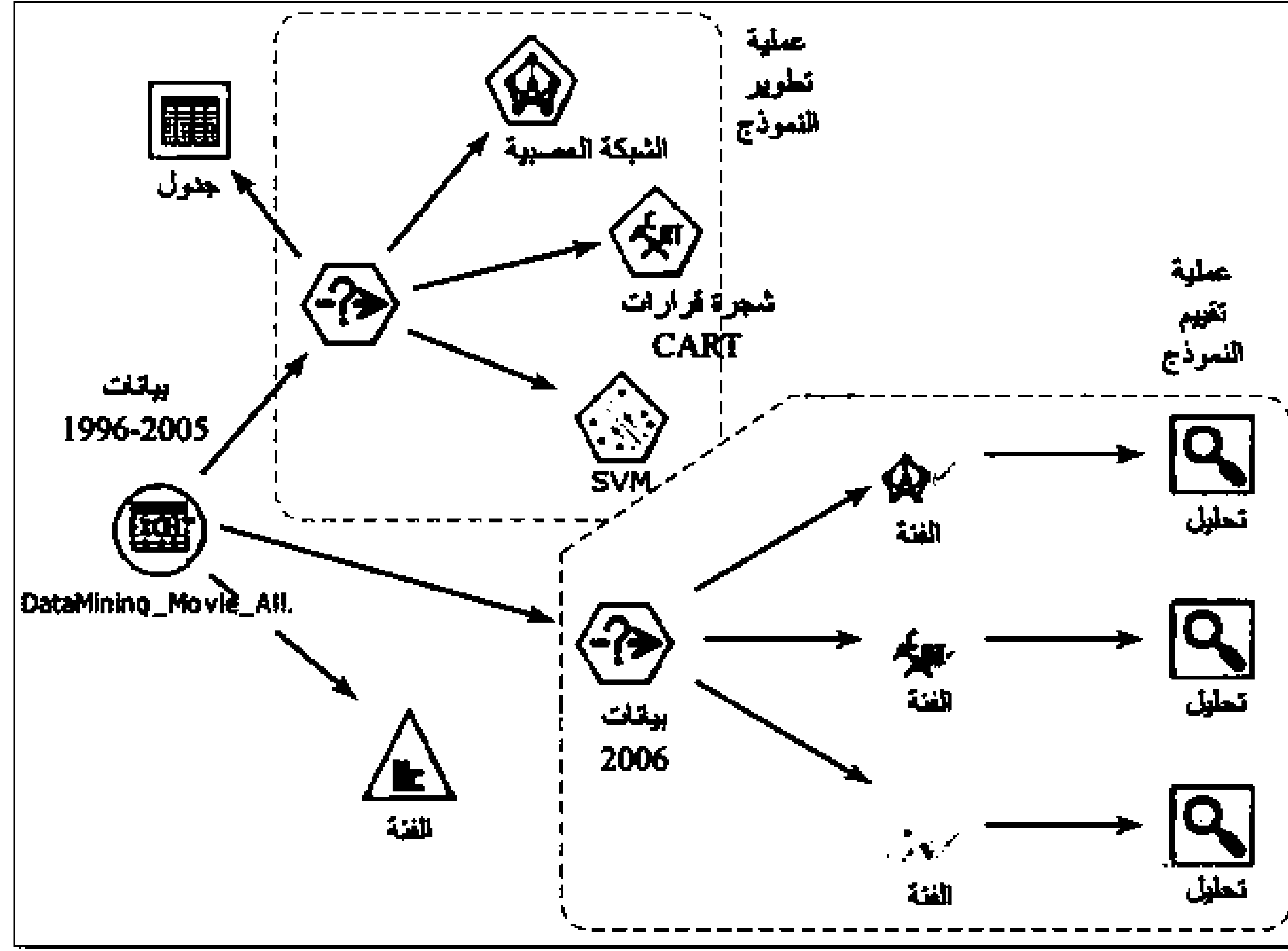
جدول ٤-٤: ملخص للمتغيرات المستقلة

المتغير المستقل	عدد القيم	القيم الممكنة
تصنيف الاتحاد الأمريكي للصور المتحركة MPAA	٥	G, PG, PG-13, R, NR
المنافسة	٣	مرتفع، متوسط، منخفض
قيمة النجم	٣	مرتفع، متوسط، منخفض
النوع الأدبي	١٠	خيــال علمــي، درامــا ملحمية تاريخيــة، دراما حديثة، سياســة، إثارة، رعــب، كوميديا، كرتون، حركي، وثائقي
مؤثرات خاصة	٣	مرتفع، متوسط، منخفض
متمم	٢	نعم، لا
عدد الشاشات	١	عدد صحيح موجب بين ١ و٣٨٧٦

المنهجية:

باســتخدام مجموعةٍ متنوعةٍ من أساليب التنقيب في البيانات، بما في ذلك الشبكات العصبية وأشجار القرار وآلة المتجهات الداعمة وثلاثة أنواع من المجموعات؛ قام كلٌّ من Sharda وDelen بتطويــر نماذج التنبؤ. وقد تمَّ اســتخدامُ بيانــات الحقبة الزمنية من ١٩٩٨ إلى ٢٠٠٥م كبيانــات تدريبيــة لبناء نماذج التنبؤ، في حين اســتُخدِمت بيانات عام ٢٠٠٦م كبيانات اختبار لتقييم ومقارنة دقة التنبؤات في النماذج. ويعرض الشكل ٤-١٦

لقطة شاشة من IBM SPSS Modeler (أداة التنقيب في البيانات Clementine سابقًا) التي تصوِّر خريطة العملية المستخدمة لمشكلة التنبؤ؛ إذ يعرض الجانب الأيسر العلوي للخريطــة عملية تطوير النموذج، ويعرض الركن الســفلي الأيمن مــن الخريطة عملية تقييم النموذج (أيْ: الاختبار أو التسجيل). ولمزيدٍ من التفاصيل حول أداة IBM SPSS Modeler وكذلك استخدامها فيُرجَى الرجوع إلى موقع الكتاب على شبكة الإنترنت).



شكل ٤-١٦: لقطة لتدفق عملية نظام التنبؤ بشباك التذاكر

النتائج:

يقــدِّم الجدول ٤-٥ نتائج التنبؤ لجميع طرق الجمع الثـلاث للبيانات بالإضافة إلى نتائــج المجموعات الثلاث المختلفة. يُعَدُّ أول قياس للأداء هو معدل التصنيف الصحيح للنســبة المئوية، ويُطلَق عليه «bingo». وكما ورد في الجدول معدل التصنيف الصحيح 1-Away (أيْ: ضمن فئة واحدة). تشير النتائج التي يعرضها الجدول إلى أن SVM كان أفضل نماذج التنبؤ الفردية، يليها ANN؛ في حين حلَّت خوارزمية شجرة قرار CART في المركز الأخير كأســوأ نموذج. بشــكلٍ عام؛ حققت نماذج المجموعة أداءً أفضل من نماذج التنبــؤ الفردية، التي أدَّت خوارزمية الاندمــاج فيها أفضل أداءٍ ممكن. ويُعَدُّ الانحراف المعيــاري المنخفض للغاية الذي تم الحصــول عليه من المجموعات هو - على الأرجح - أكثر أهمية لصانعي القرار وقراءة جدول النتائج، مقارنة بالنماذج الفردية.

جدول ٤-٥: نتائج التنبؤ المجدولة للنماذج الفردية والجماعية

		نماذج التنبؤ				
		النماذج الفردية	النماذج الجماعية			
مقياس الأداء	SVM	ANN	CART	الغابة العشوائية	الشجرة المدعومة	الاندماج (متوسط)
العد (bingo)	١٩٢	١٨٢	١٤٠	١٨٩	١٨٧	١٩٤
العد (1-Away)	١٠٤	١٢٠	١٢٦	١٢١	١٠٤	١٢٠
الدقة (٪bingo)	٪٤٩,٥٥	٪٦٠,٥٢	٪٤٠,٤٦	٪٦٢,٥٤	٪٥٤,٠٥	٪٥٦,٠٧
الدقة (٪1-Away)	٪٨٥,٥٥	٪٢٨,٨٧	٪٨٨,٧٦	٪٨٩,٦٠	٪١٠,٨٤	٪٩٠,٧٥
الانحراف المعياري	٠,٩٣	٠,٨٧	١,٠٥	٠,٧٦	٠,٨٤	٠,٦٣

الاستنتاج:

يدَّعي الباحثون أنَّ نتائج التنبؤ هذه أفضل من أيِّ تقارير منشورة فيما يتعلق بنطاق هذه المشكلة. وبالإضافة إلى الدقة الجذابة لنتائج التنبؤ بإيرادات الشباك؛ فإنه يمكن استخدام هذه النماذج كذلك لتحليل متغيرات القرار (وربما تحسينها) لتعظيم العائد المالي. وعلى وجه التحديد؛ يمكن تغيير المعلمات المُستَخدَمة للنمذجة باستخدام نماذج التنبؤ المجرَّبة بالفعل للوصول إلى فهمٍ أفضل لتأثير المعلمات المختلفة على النتائج النهائية. وخلال هذه العملية، والتي تُسمَّى عادةً «تحليل الحساسية»؛ يمكن لصانع القرار في إحدى الشركات الترفيهية أن يكتشف، بمستوى دقة عالٍ إلى حَدٍّ ما، مقدار قيمة النجاح المالي التي يمكن أن يحققها ممثل معين (أو تاريخ صدور محدد، أو إضافة المزيد التأثيرات الفنية... إلخ)؛ مما يجعل النظام الأساسي أداةً لا تُقدَّر بثمن لدعم القرار.

أسئلة للمناقشة:

١- لماذا يعتبر التنبؤ بالنجاح المالي للأفلام أمرًا مهمًّا بالنسبة للعديد من صناع السينما في هوليوود؟

٢- كيف يمكن استخدام التنقيب في البيانات للتنبؤ بالنجاح المالي للأفلام قبل بدء عملية الإنتاج؟

٣- كيف تعتقد أن هوليوود نجحت في هذه المهمة إلى الآن دون مساعدة من أدوات وتقنيات التنقيب في البيانات؟

Sources: Sharda, R., & Delen, D. (2006). Predicting box-office success of motion pictures with neural networks. Expert Systems with Applications, 30, 243254-; Delen, D., Sharda, R., & Kumar, P. (2007). Movie forecast Guru: A Web-based DSS for Hollywood managers. Decision Support Systems, 43(4), 11511170-.

أسئلة مراجعة على القسم ٤-٦:

١- ما هي أدوات التنقيب في البيانات التجارية الأكثر انتشارًا؟

٢- لماذا تعتقد أن أكثر الأدوات انتشارًا يتمُّ تطويرها من قِبل الشركات القائمة على الإحصاء؟

٣- ما هي أكثر أدوات التنقيب في البيانات المجانية انتشارًا؟ ولماذا تتمتع بهذا الانتشار الهائل (خاصة R)؟

٤- ما هي الاختلافات الرئيسة بين أدوات البرمجيات التجارية وأدوات التنقيب في البيانات المجانية؟

٥- ما هي معايير الاختيار الخمسة الأولى لأداة التنقيب في البيانات؟ اشرح إجابتك.

٤-٧ مشكلات خصوصية التنقيب في البيانات، والخرافات، والأخطاء الفادحة:

غالبًا ما تحتوي البيانات التي يتمُّ جمعها وتخزينها وتحليلها في مجال التنقيب في البيانات على معلومات حول الأشخاص الحقيقيين. وقد تتضمن هذه المعلومات ما يلي:

١- بيانات التعريف (الاسم، والعنوان، ورقم الضمان الاجتماعي، ورقم رخصة القيادة، ورقم الموظف، وما إلى ذلك).

٢- البيانات الديموغرافية (مثل: العمر، والجنس، والعرق، والحالة الاجتماعية، وعدد الأطفال).

٣- البيانات المالية (مثل: الراتب، أو دخل الأسرة الإجمالي، أو التحقق من رصيد الحساب أو التوفير، أو ملكية المنزل، أو تفاصيل حساب الرهن العقاري أو القرض، أو حدود بطاقات الائتمان والأرصدة، أو تفاصيل حساب الاستثمار).

٤- تاريخ الصفقات المُبرَمة (أيْ: ما يتم شراؤه من أين ومتى، سواء من سجلات المعاملات الخاصة بالبائع أو من المعاملات الخاصة ببطاقة الائتمان).

٥- البيانات الشخصية (مثل: ذكرى الزواج السنوية، الحمل، المرض، الوفيات في العائلة، إيداعات الإفلاس).

ويمكــن الوصــول إلى معظم هذه البيانات من خلال بعض مقدِّمــي البيانات التابعين لجهات خارجية. والســؤال الرئيس هنا هو عن خصوصية الشخص الذي تنتمي إليه البيانات. وللحفاظ عــلى خصوصية وحمايــة حقوق الأفراد؛ يجب أن يتمتع متخصصو التنقيب في البيانات بســمات أخلاقيــة فضلًا عن الالتزامات القانونية. وتتمثل إحدى طرق تحقيق ذلك في عملية إلغاء تحديد ســجلات العملاء قبل تطبيق تطبيقات التنقيب في البيانات؛ بحيث لا يمكن تتبُّع السجلات للفرد. وقد تمَّ بالفعل إلغــاء تحديد العديد من مصادر البيانات المتاحة للجمهور (مثل: بيانات CDC، بيانات SEER، بيانات UNOS). وقبل الوصول إلى مصادر البيانات هذه، يُطلب من المستخدمين في كثيرٍ من الأحيان الموافقة على أنهم - تحت أيِّ ظرف من الظروف - ســيحاولون تحديد هوية الأفراد الذين يقفون وراء هذه الأرقام.

كان هنــاك عددٌ من الحالات في الماضي القريب؛ إذ شــاركت بعض الــشركات بيانات العملاء مع الآخرين دون الحصول على موافقة صريحة من عملائها. فعلى ســبيل المثال - كما قد يتذكر معظمكم - في عام ٢٠٠٣م، قدَّمت شركة JetBlue Airlines الســجلات الخاصة بأكثر من مليون مسافر من عملائها لشركة Torch Concepts، وهي مقاولة حكومية أمريكية. ثم قامت Torch بعد ذلك بتعزيز بيانات الركاب بمعلومات إضافية، مثل: أحجام العائلة وأرقام الضمان الاجتماعي، وهي معلومات تم شراؤها من وسيط بيانات يُسمَّى Acxiom. وكان القصد من قاعدة البيانات الشــخصية الموحَّدة أن تُســتَخدَم في مشروع التنقيب في البيانات؛ بغــرض التنبؤ بملامح الإرهاب المحتمل. وقد تمَّ كلُّ هذا دون إخطار أو موافقة الركاب. وعندما خرجت أخبار تلك الأنشــطة؛ تــمَّ تقديمُ العشرات من قضايا الخصوصية ضــد JetBlue وTorch وAcxiom، ودعا العديد من أعضاء مجلس الشيوخ الأمريكي إلى إجراء تحقيق في الحادث (Wald, 2004). وخرجت في الماضي القريب أخبارٌ مماثلةٌ متعلقة بالخصوصية عن شركات الشبكات الاجتماعية الشهيرة، والتي أُشِيع أنها كانت تبيع بيانات خاصة بالعملاء إلى شركات أخرى للتسويق المستهدف الشخصي.

كانت هناك قصةٌ غريبةٌ أخرى عن مخاوف الخصوصية التي وصلت إلى عناوين الأخبار في عام ٢٠١٢م. في هذه الحالة، لم تســتخدم الشركة حتى أي بيانات خاصة و/ أو شــخصية. من الناحية القانونيــة، لم يكن هناك انتهاكٌ لأية قوانين. وتلخص الحالة العملية ٤-٧ تلك القضية - والمتعلقة بشركة Target - بشيء من التفصيل.

حالة عملية ٤-٧

توقع أنماط شراء العملاء - قصة Target

تدور القصة حول فتاة مراهقة كانت تشـارك منشورات إعلانية وكوبونات من قِبل Target لأنواع الأشياء التي تشتريها السيدة الحامل للمرة الأولى من متجر مثل Target. تقـول القصـة: ذهب رجـل غاضب إلى شركـة Target خـارج Minneapolis، مطالبًا بالتحدث إلى أحد المديرين؛ فقال: «اسـتلمت ابنتي هـذا منكم عن طريق البريد، إنها لا تزال في المدرسة الثانوية وترسلون لها كوبونات لملابس الأطفال وأسِرَّة الأطفال حديثي الولادة، هل تحاولون تشجيعها على الحمل؟». لم يكن لدى المدير أيُّ فكرة عمّا يتحدث عنه الرجل. نظر في المرسَـل؛ فوجد أنها كانت موجَّهة إلى ابنة الرجل وتضمَّنت إعلانات عن ملابس الأمومة وأثاث الحضانة وصور الرضع. اعتذر المدير ثم اتصل بعد بضعة أيام عـلى الهاتف للاعتذار مرةً أخرى. ومع ذلك؛ بدا الأب متضايقاً إلى حدٍّ ما عندما ردَّ على المدير قائلًا: «لقد تحدثت مع ابنتي وتبَيَّن لي أن هناك بعضَ الأنشطة في منزلي لم أكن على دراية بها. إنها في انتظار مولودها الأول بالفعل وأنا مدينٌ لك باعتذار». ولكن السـؤال هنـا كيف عرفت Target أن الفتاة كانت حاملًا قبل اكتشـاف والدها للحقيقة؟ وفيما يلي الإجابة عن هذا السـؤال: إن شركة Target تُخصِّص لكلِّ عميل رقم بطاقة تعريف كضيف لديها (مرتبطة ببطاقة الائتمان، أو الاسم، أو عنوان البريد الإلكتروني) الذي يصبح عنصرًا نائبًا يحتفظ بسـجلٍ لكلِّ ما اشـتراه. وتقوم Target بتعزيـز هذه البيانات بأيِّ معلومـات ديموغرافيـة قاموا بجمعها من العملاء أنفسـهم، أو تـم شراؤها من مصادر معلومات أخرى. وباسـتخدام هذه المعلومات؛ نظرت شركـة Target في بيانات الشراء السابقة لجميع الإناث اللواتي قمن بتسجيل أسمائهن في سجلات أمهات الأطفال المنتظر اسـتقبالهم. وقاموا بتحليـل البيانات من جميع الاتجاهات، وسرعـان ما ظهرت بعض الأنماط المفيدة. فعلى سـبيل المثال: كانت مستحضرات اللوشـن والفيتامينات الخاصة من بين المنتجات ذات أنماط الشراء المثيرة للانتباه. فالكثيرُ من الناس يشـترون اللوشـن، ولكـن ما لاحظوه هو أن السـجلات التي تحتـوي على بيانات سـيدات ينتظرن أطفال يقمن بشراء كميات أكبر من محلول اللوشـن غير المعطر مع بداية الثلث الثاني من فترة الحمل. ولاحظ محللٌ آخر أنه في وقتٍ ما في الأسـابيع العشرين الأولى من فترة الحمل، تقوم النسـاء الحوامل بالاعتماد على المكملات مثل: الكالسيوم والمغنيسيوم والزنك. كما يقوم العديدُ من المتسوِّقين بشراء صابون وكرات قطنية، ولكن عندما تبدأ سيدة ما فجأة

بشراء الكثير من الصابون الخالي من الرائحة وأكياس كبيرة من كرات القطن، بالإضافة إلى مطهرات اليد ومناديل الحمام؛ فإن ذلك يشـير إلى قرب موعد اسـتقبالها لمولودها. وفي النهايـة؛ تمكّنوا من تحديد حوالي ٢٥ منتجًـا، عندما تم تحليلها معًا، مكّنتهم من تحديد درجة توقع وجود حمل من عدمه لكل متسـوق. والأهم من ذلك؛ اسـتطاعت Target أيضًا تقدير تاريخ الولادة لكلِّ سـيدة حامل؛ لذلك يتم إرسـال كوبونات منتجات معينة بتواريخ مُحدَّدة لكلِّ متسوقة بحسب المرحلة التي تمرُّ بها في حملها.

إذا نظرت إلى هذه الممارسـة من منظور قانوني؛ فستستنتج أن Target لم تستخدم أيَّ معلومـات تنتهـك خصوصية العميل؛ إنهم فقط اسـتخدموا بيانات المعاملات التي تقوم كلُّ سلسـلة تجارة التجزئة الأخرى تقريبًا بجمعها وتخزينها (وربما تحليلها) حول عملائها. ولكن ما كان مزعجًا في هذا السـيناريو ربما كان المفهوم المسـتهدف: الحمل. فهنـاك بعض الأحداث أو المفاهيم يجب أن تكون غير قابلة للتداول أو أن يتم التعامل معها بحذر شديد، مثل: أمراض الشيخوخة، والطلاق، والإفلاس.

أسئلة للمناقشة:

١- مـا رأيك في التنقيـب في البيانات وتأثيرهـا على الخصوصية؟ ومـا هي العتبة الفاصلة بين اكتشاف المعرفة وانتهاك الخصوصية؟

٢- هل أخطأت Target؟ هل فعلت أيَّ شيء غير قانوني؟ برأيك، ما الذي كان يجب أن تفعله Target؟ وهل يجب على Target الإقلاع عن هذه الأنواع من الممارسات مستقبلًا؟

Sources: Hill, K. (2012, February 16). How Target figured out a teen girl was pregnant before her father did. Forbes; Nolan, R. (2012, February 21). Behind the cover story: How much does Target know? NYTimes.com.

خرافات وأخطاء فادحة حول التنقيب في البيانات:

يُعَـدُّ التنقيبُ في البيانات أداةً تحليليةً قويـةً تُمكِّن المديرين التنفيذيين من التقدُّم في وصف طبيعة الماضي (بالنظر إلى مرآة الرؤية الخلفية) للتنبؤ بالمستقبل (النظر إلى الأمام) لإدارة عملياتهم التجارية بشـكلٍ أفضل (اتخاذ قرارات دقيقة وفي الوقت المناسـب). تسـاعد عملية التنقيب في البيانات جهات التسـويق في العثـور على أنماطٍ تفصح عن أسرار تتعلّق بسـلوك العملاء. يُمكِن اسـتخدام نتائج التنقيب في البيانات لزيادة الإيرادات وخفض التكلفة من خلال تحديد الاحتيال واكتشاف الفرص التجارية؛ مما يوفر مناخًا جديدًا تمامًا وجيدًا للمَيْزة التنافسية. وكحقل متطور

وناضج، يرتبط التنقيب في البيانات غالبًا بعددٍ من الخرافات، والتي يعرض الجدول ٤-٦ بعضًا منها (Delen, 2014; Zaima, 2003).

جدول ٤-٦: خرافات التنقيب في البيانات

الخرافة	الواقع
يوفر التنقيب في البيانات تنبؤات تشبه كرة بلورية فورية.	تُعَدُّ عملية التنقيب في البيانات عمليةً متعددة الخطوات تتطلب تصميمًا واستخدامًا متعمدًا واستباقيًا.
التنقيب في البيانات غير قابلٍ للتطبيق مع تطبيقات الأعمال السائدة.	الحالة الحالية للتنقيب في البيانات جاهزةٌ للاستعمال تقريبًا لأيِّ نوع أو حجم عمل.
يتطلب التنقيبُ في البيانات قاعدة بيانات منفصلة ومُخصَّصة.	بسبب التقدُّم في تقنية قاعدة البيانات؛ لا يلزم وجود قاعدة بيانات مخصصة.
يمكن فقط للأشخاص ذوي الدرجات العلمية المتقدِّمة القيام بالتنقيب في البيانات.	تمكِّن الأدوات الحديثة القائمة على الويب المديرين من جميع المستويات التعليمية من القيام بعملية التنقيب في البيانات.
التنقيب في البيانات هو فقط للشركات الكبيرة التي لديها الكثير من بيانات العملاء.	إذا كانت البيانات تعكسُ بدقة الأعمال أو عملائها، يمكن لأيِّ شركة استخدام التنقيب في البيانات.

لقد اكتسب مناصرو التنقيب في البيانات فائدةً تنافسيةً هائلةً من خلال فهم أنَّ هذه الخرافات هي مجرد: خرافات. وعلى الرغم من أن قيمة الاقتراح وبالتالي ضرورة وجودها واضحة لأيِّ شخص؛ فإن أولئك الذين ينفذون مشاريع التنقيب في البيانات سواءً المبتدئ أو عالم البيانات المتمرس يرتكبون أخطاءً أحيانًا تؤدي إلى مشاريع ذات نتائج أقل من المرغوب فيها. ونعرض فيما يلي ١٦ من الأخطاء الشائعة في التنقيب في البيانات، ويجب أن يكون علماء البيانات على دراية بها. وعليه، يتحتم عليهم بذل قصارى جهدهم لتجنبها Nesbit وآخرون Shultz, 2004; Skalak, 2001 2009:

١- اختيار مشكلة خطأ للتنقيب في البيانات: فلا يمكن حلُّ كلِّ مشكلة تجارية بالتنقيب في البيانات (أيْ: متلازمة الطلقة السحرية). وعندما لا توجد بيانات تمثيلية (كبيرة وغنية بالميزات)، فلا يمكن أن يكون هناك مشروع استخراج بيانات عملي.

٢- تجاهل مايعتقده الداعمون في ما يمكن فعله وما لا يمكن فعله باستخدام التنقيب في البيانات: فإدارة التوقعات هي مفتاح نجاح مشاريع التنقيب في البيانات.

٣- البـدء دون وَضْـع تصوُّر للنهاية: عـلى الرغم من أن التنقيب في البيانات هو عملية اكتشـاف للمعرفة؛ فإنه يجب أن يكون لدى المسـتخدِم هدفٌ (مشكلة أعمال محددة) في ذهنه للنجاح. لأنه، وكما يقول المثل: «إذا كنتَ لا تعرف إلى أين أنت ذاهب؛ فلن تصل إلى هناك أبدًا».

٤- تحديد المشروع الذي لا تدعمه بياناتك: يُعَدُّ التنقيب في البيانات هو كلُّ شيء عن البيانات؛ بمعنى أن القيـد الأكبر لديك في مشروع التنقيـب في البيانات هو ثراءُ البيانات. فمعرفة حدود البيانات تساعدك على صياغة مشاريع قابلة للتطبيق وتحقق النتائج المرجوة منها وتفي بالتوقعات.

٥- تحديـد وقتٍ غـير كافٍ لإعداد البيانات: إذ يتطلَّب التجهيزُ مجهـودًا أكبر من المتوقع غالبًا. حيث تشـير المعرفة الشائعة إلى أن ما يصل إلى ثلث إجمالي وقت المشروع يتمُّ استهلاكه على مهـامّ جمع البيانات، وفهمها، وتجهيزها. ولتحقيق النجاح؛ تجنَّب البدء في النمذجة قبل أن تتم معالجة بياناتك بشكل صحيح (بحيث تكون متكاملة ومُنظَّفة ومُحَوَّلة).

٦- النظـر فقـط في النتائج المجمَّعة وليس في السـجلات الفردية: يكـون التنقيب في البيانات في أفضل حالاته عندما يتمُّ تمثيل البيانات بشكلٍ دقيق. حاول تجنُّب تجميع البيانات وتبسيطها بشـكلٍ لا داعي له لمسـاعدة خوارزميات التنقيب في البيانات؛ فهي لا تحتاج إلى مسـاعدتك حقًّا؛ فلديها قدرة فائقة على فعل كل ذلك بنفسها.

٧- الوقوع في فخِّ تتبُّع إجراءات ونتائج التنقيب: ولأنها عمليةُ اكتشـافٍ تنطوي على العديد من التكـرارات والتجارب؛ فمن المرجَّح أن تفقد مسـار النتائج. يتطلب النجاح تخطيطًا ممنهجًا ومنظمًا وتنفيذًا وتتبُّعًا / تسجيلًا لجميع مهام التنقيب في البيانات.

٨- استخدام بيانات مستقبلية للتنبؤ بالمستقبل: بسبب عدم وجود وصف وفهم للبيانات، في كثير مـن الأحيان يقوم مُحلِّلو البيانات بتضمين متغيرات غير معروفة في الوقت الذي من المفترض أن يتمَّ التنبؤ فيه. ومن خلال القيام بذلك؛ تنتج نماذج التنبؤ الخاصة بها نتائج دقيقة لا يمكن تصديقها (وهي ظاهرة يُطلق عليها غالبًا «ذهب الأغبياء»). وإذا كانت نتائج التنبؤ الخاصة بك جيدة جدًا لدرجة يصعب تصديقها؛ فإنها عادةً ما تكون كذلك؛ وفي هذه الحالة، فإن أول شيء تحتاج للبحث عنه هو الاستخدام غير الصحيح لمتغير من المستقبل.

٩- تجاهل النتائج المشـبوهة وانتقل بعيدًا عنها بسرعة: غالبًا ما تكون النتائج غير المتوقعة هي مؤشرات المسـتجدات الحقيقية في مشـاريع التنقيب في البيانات. ويمكن أن يؤدي التحقيق السليم في مثل هذه النتائج الشاذة إلى اكتشافات مثيرة للدهشة.

١٠- البـدء بمـشروع مجمـع رفيع المسـتوى مـن شـأنه أن يجعلك نجمًا فـوق العـادةً: غالبًا ما تفشل مشاريع التنقيب في البيانات إذا لم يتم التفكير فيها بعناية من البداية إلى النهاية.

فغالبًا ما يأتي النجاح بتقدُّم منتظم ومنظم للمشاريع من الصغيرة / البسيطة إلى الكبيرة / المعقدة. ويجب أن يكون الهدف هو إظهار القيمة المضافة التدريجية والمستمرة، بدلًا من الدخول في مشروعٍ كبيرٍ يستهلك الموارد دون إنتاج أيِّ نتائج ذات قيمة.

١١- تشغيل خوارزميات التنقيب في البيانات بشكلٍ متكرِّر وأعمى: على الرغم من أن أدوات التنقيب في البيانات اليوم قادرةٌ على استهلاك البيانات وإعداد معلمات خوارزمية لتحقيق نتائج؛ فإنه ينبغي لمستخدِم البيانات معرفة كيفية تحويل البيانات وتعيين قيم المعلمات المناسبة للحصول على أفضل النتائج الممكنة. فلكلِّ خوارزمية طريقةٌ فريدةٌ خاصةٌ بها لمعالجة البيانات، ومعرفة ما هو ضروري للحصول على أقصى استفادة من كلِّ نوع من أنواع النماذج.

١٢- تجاهل خبراء الموضوع: يتطلَّب فهم نطاق المشكلة والبيانات ذات الصِّلة تعاونًا شديدًا بين التنقيب في البيانات وخبراء المجال. فالعمل الجماعي يساعد خبير التنقيب في البيانات على تجاوز التمثيل النحوي والحصول أيضًا على الطبيعة الدلالية (أيْ: المعنى الحقيقي للمتغيرات) للبيانات.

١٣- تصديق كلِّ ما قيل لك عن البيانات: على الرغم من ضرورة التحدُّث إلى خبراء المجال لفهم البيانات ومشكلة العمل بشكلٍ أفضل؛ فإنَّ عالم البيانات يجب ألا يأخذ أيَّ شيء على أنه أمرٌ مُسلَّمٌ به. إن الفاعلية والتحقق من خلال التحليل النقدي هما مفتاح الفهم الحميم للبيانات ومعالجتها.

١٤- افتراض أن حُماة البيانات سيتواجدون بشكلٍ كاملٍ للتعاون في أيِّ وقتٍ يُطلَب منهم ذلك: تفشل العديدُ من مشاريع التنقيب في البيانات؛ لأن خبير التنقيب في البيانات لا يعرف/ لا يفهم السياسات التنظيمية. قد يكون أحد أكبر العقبات في مشاريع التنقيب في البيانات هم الأشخاص الذين يملكون البيانات ويسيطرون عليها. إن فهم وإدارة السياسة هو مفتاح تحديد البيانات والوصول إليها، وفهمها بشكلٍ صحيح لإنتاج مشروع ناجح للتنقيب في البيانات.

١٥- اختلاف معيار قياس النتائج بين المحلل وصاحب البيانات: يجب أن توجَّه النتائج إلى المستخدِم النهائي (المدير / صانع القرار) الذي سيستخدمها. وبالتالي؛ فإن تقديم النتائج في مقياسٍ وشكلٍ يجذب المستخدِم النهائي يزيد بشكلٍ كبير من احتمالية الفهم الحقيقي والاستخدام السليم لنتائج التنقيب في البيانات.

١٦- إذا قمت ببنائه؛ فسيأتي: لا تقلق بشأن كيفية تقديمه: عادةً، يعتقد خبراء التنقيب في البيانات أنها تتم بمجرد إنشاء نماذج جيدة تلبِّي وتزيد على احتياجات / متطلبات / توقعات

المستخدِم النهائي (أيْ: العميل). وبدون انتشار جيد؛ تكون القيمة المُحدَّدة لنتائج التنقيب في البيانات محدودةً نوعًا ما. لذلك؛ يُعَدُّ الانتشـــار خطوةً أخيرةً ضروريةً في عملية التنقيب في البيانـــات؛ إذ يتمُّ دَمْج النماذج في البنية التحتيـــة لدعم القرار التنظيمي؛ من أجل تمكين عملية اتخاذ القرار بشكلٍ أفضل وأسرع.

أسئلة مراجعة على القسم ٤-٧:

١- ما هي مشكلات الخصوصية في التنقيب في البيانات؟

٢- كيف تعتقد أن يتحسَّن النقاش بين الخصوصية والتنقيب في البيانات؟ ولماذا؟

٣- ما هي أكثر الخرافات شيوعًا حول التنقيب في البيانات؟

٤- ما رأيك في أسباب هذه الخرافات حول التنقيب في البيانات؟

٥- ما هي الأخطاء الأكثر شيوعًا في التنقيب في البيانات؟ وكيف يمكن تخفيفها أو القضاء عليها تمامًا؟

تلخيص لأهمِّ نقاط الفصل:

- يُعَدُّ التنقيبُ في البيانات عمليةً لاكتشاف المعرفة الجديدة من قواعد البيانات.
- يُمكِن أن يستخدم التنقيب في البيانات ملفات مسطحة بسيطة كمصادر بيانات أو يمكن تنفيذها على بيانات في مستودعات البيانات.
- هناك العديدُ من الأسماء والتعريفات البديلة للتنقيب في البيانات.
- يتواجـــد التنقيـــب في البيانات في تقاطـــع العديد من التخصصات، بما في ذلـــك الإحصاء والذكاء الاصطناعي والنمذجة الرياضية.
- تستخدم الشركات التنقيب في البيانات؛ لفهم عملائها بشكلٍ أفضل وتحسين عملياتها.
- يمكن العثور على تطبيقات التنقيب في البيانات في كلِّ مجال من مجالات الأعمال والحكومة، بما في ذلك الرعاية الصحية والمالية والتسويق والأمن الداخلي.
- يُعَـــدُّ كلٌّ من التنبـــؤ (التصنيف أو الانحدار)، والتجميع، والترابط ثلاث فئاتٍ واســـعة من مهام التنقيب في البيانات.
- مثل مبادرات التعزيز المؤسسي الأخرى؛ يجب أن يتبع مشروع استخراج البيانات عملية منهجية لإدارة المشاريع لتكون ناجحة.

- تمَّ اقتراحُ العديد من عمليات التنقيب في البيانات: CRISP-DM، وSEMMA، وKDD وما إلى ذلك.
- يوفِّر CRISP-DM طريقةً منتظمةً ومنظمةً لإجراء مشاريع التنقيب في البيانات.
- تســتهلك الخطوات الســابقة في مشــاريع التنقيب في البيانات (أيْ: فهم المجال والبيانات ذات الصِّلة) معظم الوقت الإجمالي للمشروع (غالبًا ما يزيد عن ٨٠٪ من الوقت الإجمالي).
- تُعَــدُّ المعالجة الأوَّلية للبيانات أمرًا ضروريًّا لأيِّ دراســة ناجحــة للتنقيب في البيانات؛ إذ تؤدي البيانات الجيدة إلى معلومات جيدة، والمعلومات الجيدة تؤدي إلى قرارات جيدة.
- تتضمَّن معالجة البيانات الأوَّلية أربع خطوات رئيسة: دَمْج البيانات، وتنظيف البيانات، وتحويل البيانات، وتقليل حجم البيانات.
- تتعلَّم مناهج التصنيف من الأمثلة السابقة التي تحتوي على مدخلات وعلامات الفئات الناتجة، وبمجرد أن يتمَّ تدريبهم بشكلٍ صحيح؛ يكونون قادرين على تصنيف الحالات المستقبلية.
- يتمُّ تجميعُ ســجلات نمط الأقســام إلى شرائح أو مجموعات طبيعية؛ بحيث يتشــابه أعضاء كل شريحة في خصائص معينة.
- يتمُّ استخدام عددٍ من الخوارزميات المختلفة بشكلٍ شائعٍ بغرض التصنيف. وتشمل التطبيقات التجارية ID3، وC4.5، وC5، وCART، وCHAID، وSPRINT.
- تقوم أشــجار القرار بتقســيم البيانات عن طريق التفرُّع عبر سمات مختلفة؛ بحيث يكون لكلِّ مجموعة أوراق نمط فئوي مشترك.
- يُعَدُّ مؤشر Gini والحصول على المعلومات (الانتروبيا)؛ طريقتين شائعتين لتحديد خيارات التفرُّع في شجرة القرار.
- يقيــس مؤشرُ Gini نقاء العينــة. فإذا كان كلُّ شيء في عينة ينتمي إلى فئة واحدة، تكون قيمة مؤشر Gini صفر.
- تستطيع عِدة تقنيات تقييم أن تقيس دقة التنبؤ لنماذج التصنيف، بما في ذلك التقسيم البسيط، التحقق المتقاطع k-fold، bootstrapping، والمساحة تحت منحنى ROC.
- يتمُّ اســتخدام خوارزميات التجميع عندما لا تحتوي ســجلات البيانات على معرَّفات فئة معرفة مسبقًا (بمعنى، لا يُعرف إلى الفئة التي ينتمي إليها سجل معين).
- تحسب خوارزميات التجميع مقاييس التشابه؛ من أجل تجميع حالات مماثلة في مجموعات.
- يُعَدُّ قياس المسافة هو مقياس التشابه الأكثر استخدامًا في التحليل التجميعي.
- يُعَدُّ كلٌّ من K-means والخرائط ذاتية التنظيم؛ هما أكثر خوارزميات التجميع شيوعًا.

- يُستخدَم التنقيب في قواعد الارتباط؛ لاكتشاف اثنين أو أكثر من العناصر (أو الأحداث أو المفاهيم) التي تذهب معاً.
- يُشار عادةً إلى التنقيب في قواعد الارتباط بمصطلح تحليل سلة السوق.
- إنَّ خوارزمية الارتباط الأكثر استخدامًا هي Apriori؛ إذ يتمُّ تحديد العناصر المتكرِّرة من خلال اتباع نهجٍ تصاعديٍّ (من أسفل إلى أعلى).
- يتمُّ تقييم قواعد الارتباط بناءً على مقاييس الدَّعم والثقة المتعلقَيْن بها.
- تتوفر العديد من أدوات التنقيب في البيانات سواء كانت مجانية أو بمقابل مادي.
- إن أدوات التنقيب في البيانات التجارية الأكثر شعبية هي SPSSPASW، وSAS Enterprise Miner.
- إن أشهر أدوات التنقيب في البيانات المجانية هي Weka وRapidMiner.

مصطلحات أساسية:

خوارزمية Apriori	شجرة القرار	اكتشاف المعرفة في قواعد البيانات (KDD)	تنبؤ
المنطقة تحت منحنى ROC	قياس المسافة	مصعد	RapidMiner
ارتباط	طاقم	تحليل الارتباط	انحدار
Bootstrapping	إنتروبيا	Microsoft Enterprise	SEMMA
بيانات تسلسلية	مؤشر Gini	التحالف	تسلسل التنقيب
تصنيف	كسب المعلومات	خادم مايكروسوفت SQL	التقسيم البسيط
تجمع	بيانات فاصلة	بيانات اسمية	الدعم
الثقة	التحقق بـ س / جزء	بيانات رقمية	Weka
CRISP-DM	KNIME	بيانات ترتيبية	التنقيب في البيانات

أسئلة للمناقشة:

١- عرِّف التنقيب في البيانات. لماذا هناك العديدُ من الأسماء والتعريفات للتنقيب في البيانات؟

٢- ما هي الأسباب الرئيسة وراء الرّواج الحديث للتنقيب في البيانات؟

٣- ناقش ما يجب على المنظمة مراعاته قبل اتخاذ قرارٍ بشراء برنامج التنقيب في البيانات.

٤- قارن بين التنقيب في البيانات والأدوات والتقنيات التحليلية الأخرى.

٥- ناقش طرق التنقيب في البيانات الرئيسة. وما هي الاختلافات الأساسية فيما بينها؟

٦- ما هي مجالات التطبيق الرئيسة للتنقيب في البيانات؟ ناقش القواسم المشتركة لهذه المجالات التي تجعلها فرصةً لدراسات التنقيب في البيانات.

٧- لماذا نحتاجُ إلى عملية مُوحَّدة للتنقيب في البيانات؟ وما هي عمليات التنقيب في البيانات الأكثر شيوعًا؟

٨- ناقش الاختلافات بين عمليتَيْ التنقيب في البيانات الأكثر استخدامًا.

٩- هل تُعَدُّ عمليات التنقيب في البيانات مجرد مجموعة متسلسلة من الأنشطة؟ اشرح إجابتك.

١٠- لماذا نحتاج إلى المعالجة المسبقة للبيانات؟ وما هي المهامُّ الرئيسة والتقنيات ذات الصلة المُستَخدَمة في المعالجة المسبقة للبيانات؟

١١- ناقش الأسباب الكامنة وراء تقييم نماذج التصنيف.

١٢- ما هو الفرق الرئيس بين التصنيف والتجميع؟ اشرح إجابتك باستخدام أمثلة من الواقع.

١٣- بالانتقال إلى ما بعد مناقشة هذا الفصل، أين يمكن استخدام الارتباط؟

١٤- ما هي مشكلات الخصوصية في التنقيب في البيانات؟ هل تعتقد أنها موثقة؟

١٥- ما هي الخرافات المتداولة والأخطاء الأكثر شيوعًا حول التنقيب في البيانات؟

تمارين: شبكة جامعة تيراداتا (TUN) وغيرها من التمارين اليدوية:

١- قم بزيارة teradatauniversitynetwork.com. حَدِّد دراسات الحالة والأوراق البيضاء حول التنقيب في البيانات. وقم بوصف التطورات الحديثة في مجال التنقيب في البيانات والنمذجة التنبؤية.

٢- انتقل إلى teradatauniversitynetwork.com. حَدِّد موقع ندوات الويب المتعلقة بالتنقيب في البيانات. على وجه الخصوص، حَدِّد موقع حلقة دراسية قدَّمها كلٌّ من C. Imhoff وT. Zouqes. شاهد ندوة الويب. ثم أجب عن الأسئلة التالية:

١- اذكر بعض تطبيقات التنقيب في البيانات المثيرة للاهتمام؟

٢- ما هي أنواع المكاسب والتكاليف التي يمكن أن تتوقعها المنظمات من مبادرات التنقيب في البيانات؟

٣- بالنسبة لهذا التمرين، يتمثل هدفك في بناء نموذج لتحديد المدخلات أو المتنبئات التي تميِّز العملاء المحفوفين بالمخاطر من الآخرين (بناءً على الأنماط المتعلقة بالعملاء السابقين) ثم استخدام هذه المدخلات للتنبؤ بعملاء جدد محفوفين بالمخاطر. تُعَدُّ هذه العينة نموذجيةً لهذا المجال. وبيانات النموذج المراد استخدامه في هذا التمرين هي في ملف Online W4.1 في ملف CreditRisk.xlsx. تتضمَّن مجموعة البيانات ٤٢٥ حالة و١٥ متغيرًا متعلقًا بالعملاء السابقين والحاليين الذين اقترضوا من أحد البنوك لأسباب مختلفة. كما تحتوي على معلومات متعلقة بالعميل مثل الوضع المالي، وسبب القرض، والتوظيف، والمعلومات الديموغرافية، والنتيجة أو المتغير التابع للمركز الائتماني، مع تصنيف كلِّ حالة على أنها جيدة أو سيئة، بناءً على تجربة المؤسسة السابقة. خُذ ٤٠٠ من الحالات كحالات تدريب وقم بتخصيص ٢٥ حالة أخرى للاختبار. قم ببناء نموذج شجرة قرارات لمعرفة خصائص المشكلة. اختبر أداء هذا النموذج على ٢٥ حالة أخرى. ثم قم بإعداد تقريرين، أحدهما عن أداء التعلم والاختبار في نموذجك. والآخر لتحديد نموذج شجرة القرار ومعلمات التدريب، بالإضافة إلى الأداء الناتج في مجموعة الاختبار. يمكنك استخدام أيٍّ برنامج لشجرة القرارات. (This exercise is courtesyof StatSoft, Inc., based on a German data set from ftp.ics.uc,i.edu/ pub/machine-learning-databases / يهدف هذا التمرين إلى تعريفك بتعدين القاعدة. (statlog/german renamed CreditRisk and altered).

٤- بالنسبة لهذا التمرين، ستقوم بتكرار (على نطاق أصغر) نمذجة شباك التذاكر الموضَّح في الحالة العملية ٤-٦. قم بتنزيل مجموعة بيانات التدريب من كلٍّ من Online File W4.2، وMovieTrain.xlsx، وهي بتنسيق Microsoft Excel. استخدم وصف البيانات الموضَّح في الحالة العملية ٤-٦ لفهم المجال والمشكلة التي تحاول حلها. قم بانتقاء واختيار المتغيرات المستقلة الخاصة بك. قم بتطوير ثلاثة نماذج تصنيف على الأقل (على سبيل المثال، شجرة القرارات، الانحدار اللوجستي، الشبكات العصبية). قارن نتائج الدقة باستخدام تقنيات 10-fold cross-validation والتقسيم المئوي واستخدم كذلك مصفوفات الدقة ثم علِّق على النتيجة. اختبر النماذج التي قمت بتطويرها على مجموعة الاختبار (انظر ملف الإنترنت W4.3, MovieTest.xlsx). قم بتحليل النتائج بنماذج مختلفة، مع وضع أفضل نموذج تصنيف، ودعمه مع نتائجك.

٥- يهــدف هــذا التمرين إلى تقديمــك إلى التنقيب في قواعد الارتباط. تحتــوي مجموعة بيانات Excel baskets1ntrans.xlsx على ٢٨٠٠ ملاحظة / سـجل لبيانات منتجات السوبر ماركت. يحتوي كل سـجل عــلى بطاقات تعريف للعمــلاء والمنتجات التي قاموا بشرائها. اســتخدم مجموعــة البيانات هذه لفهم العلاقات بين المنتجــات (أيْ: المنتجات التي يتمُّ شراؤها معًا). ابحث عن العلاقات المثيرة، وأضِف لقطات شاشة لأيٍّ أنماط ارتباط خفية قد تجدها. وبشكلٍ أكثر تحديدًا، أجب عن الأسئلة التالية:

١- ما هي قواعد الارتباط التي تعتقد أنها الأكثر أهمية؟

٢- اســتنادًا إلى بعض قواعد الارتباط التي وجدتها، قم بإجراء ثلاث توصيات أعمال على الأقل قــد تكون مفيدة للشركة. قــد تتضمَّن هذه التوصيات أفكارًا حــول تنظيم الرَّف أو بيع المنتجات. (سيتمُّ مَنْح نقاط مكافأة للأفكار الجديدة / المبتكرة).

٣- ما هي قيم الدَّعم والثقة والرفع للقاعدة التالية؟

Wine, Canned Veg → Frozen Meal

٦- في هذه المهمة، ستستخدم أداة تعدين البيانات المجانية / مفتوحة المصدر KNIME (knime.org)؛ وذلك لإنشــاء نماذج تنبؤيــة لمجموعة بيانات تحليل حركات العملاء الصغيرة نســبيًّا. يجب عليك تحليل مجموعة البيانات المعطاة (حول ســلوك اســتبقاء / تسرُّب العميل لعينة مكونة مــن ١٠٠٠ عميــل) بغرض تطوير ومقارنة ما لا يقل عن ثلاثة نمــاذج للتنبؤ (أيْ: للتصنيف). فعلى سبيل المثال: يمكنك تضمين مقارنتك على أشجار القرار والشبكات العصبية وآلة المتجهات الداعمة والجار الأقرب و / أو نماذج الانحدار اللوجستي. فيما يلي تفاصيل هذه المهمة:

- تثبيت واستخدام أداة برنامج KNIME من (knime.org).
- يمكنك أيضًا استخدام MS Excel لتجهيز البيانات مسبقًا (حسب حاجتك / رغبتك).
- قم بتنزيل ملف البيانات CustomerChurnData.csv من موقع الكتاب على الويب.
- يتمُّ إعطاء البيانات بتنســيق CSV (قيم مفصولة بفواصل). وهذا التنســيق هو تنســيق الملفات المســطَّحة الأكثر شــيوعًا والذي يُمكِن للعديد من أدوات البرنامج فتحه / التعامل معه بسهولة (بما في ذلك KNIME وMS Excel).
- اعرض نتائجك في مستند احترافي منظم جيدًا.
- اعرض على صفحة الغلاف معلومات مناسبة عنك وعن المهمة.

- تأكَّد من تكامل الأرقام (المخططات البيانية، الرسوم البيانية، الجداول، لقطات الشاشة) بشكلٍ رائع في الوَصْف النَّصي بطريقة احترافية. كما يجب أن يحتوي التقرير على ستة أقسام رئيسية (تشبه مراحل CRISP-DM).
- حاول ألا تتجاوز ١٥ صفحة في الإجمالي، بما في ذلك الغلاف (استخدِم الخط Times New Roman حجم ١٢، وتباعد الأسطر ١,٥).

مهامُّ الفريق ومشروعات لعب الأدوار:

١- قم بفحص كيفية قيام أجهزة التقاط البيانات الجديدة، مثل علامات RFID بمساعدة المؤسسات على تحديد عملائها بدقة وتقسيمهم إلى أنشطة، مثل التسويق المستهدَف. تنطوي العديدُ من هذه التطبيقات على التنقيب في البيانات. قم بإدخال المطبوعات والويب إلى الشاشة باستخدام الماسح الضوئي (scanner)، ثم قم باقتراح خمسة تطبيقات محتملة جديدة للتنقيب في البيانات يمكنها استخدام البيانات التي تم إنشاؤها باستخدام تقنية RFID. ما هي القضايا التي يمكن أن تنشأ إذا كانت قوانين البلد تتطلب تضمين هذه الأجهزة في جسم كلِّ شخص لنظام تحديد الهوية الوطني؟

٢- قم بمقابلة المسؤولين في كليتك أو المسؤولين التنفيذيين في مؤسستك؛ لتحديد كيف يمكن أن يساعدهم التنقيب في البيانات ومستودعات البيانات وأدوات OLAP وأدوات التصوير في عملهم. اكتب اقتراحًا يَصِف نتائجك. ثم قم بإعداد تقرير يتضمَّن تقديرات التكلفة والفوائد.

٣- يتوفر مستودع جيد جدًا للبيانات التي تمَّ استخدامها لاختبار أداء العديد من خوارزميات التنقيب في البيانات في ics.uci.edu/mlearn/MLRepository.html. تهدف بعض مجموعات البيانات إلى اختبار حدود الخوارزميات الحديثة لتعلُّم الآلة، وكذلك تهدف إلى مقارنة أدائها مع الأساليب الجديدة للتعلم. ومع ذلك؛ قد تكون بعض مجموعات البيانات الأصغر مفيدةً في استكشاف وظائف أيِّ برنامج من برامج التنقيب في البيانات، مثل RapidMiner أو KNIME. قم بتنزيل مجموعة بيانات واحدة على الأقل من هذا المستودع (ولتكن، قواعد بيانات فحص الائتمان أو قاعدة بيانات الإسكان) وتطبيق أساليب شجرة القرار أو طرق التجميع حسب الاقتضاء. قم بإعداد تقرير بناءً على نتائجك. (يمكن استخدام بعض هذه التمارين، خاصةً تلك التي تنطوي على بيانات كبيرة أو مشكلات صعبة كمشاريع على المدى الطويل).

٤- يتمُّ توفير مجموعات بيانات كبيرة وغنية بالمَيِّزات من قِبل الحكومة الأمريكية أو الشركات التابعة لها على الإنترنت. فعلى سبيل المثال: يمكنك الاطلاع على مجموعة كبيرة من

مجموعات البيانات الحكومية (data.gov)، ومجموعات بيانات مراكز السيطرة على الأمراض والوقاية منها (www.cdc.gov/DataStatistics)، والمراقبة، ومجموعة البيانات الخاصة بعلم الأوبئة والنتائج النهائيـة الخاصة بشركـة (http://seer.cancer.gov/data) Cancer.org، ومجموعـات بيانات نظام تقارير تحليلات معدل الوفيات التابع لوزارة النقل (www.nhtsa.gov/FARS). لاحـظ أيضًا أنه لا يتم عمل معالجة مُسـبَقة لهـذه المجموعات من البيانات لأغـراض التنقيـب في البيانات؛ مما يجعلها مـوردًا رائعًا لتجربة عمليـة التنقيب في البيانات بالكامـل. يتم إدراج مصدر آخـر غني لمجموعة من مجموعات بيانات التحليلات على موقع KDnuggets.com (kdnuggets.com/datasets/index.html).

٥- خـذ بعين الاعتبـار مجموعة البيانات التاليـة، والتي تتضمَّن ثلاث سـمات وتصنيفًا لقرارات القبول في برنامج ماجستير إدارة الأعمال:

١- باسـتخدام البيانات المعروضة في الجدول أدناه؛ قم بتطوير قواعد الخبرة اليدوية الخاصة بك لاتخاذ القرار.

٢- اسـتخدم مؤشر Gini لإنشاء شجرة قرار. يمكنك اسـتخدام الحسابات اليدوية أو جدول البيانات أدناه لإجراء الحسابات الأساسية.

٣- استخدام برنامج شجرة قرارات آلي لإنشاء شجرة لنفس البيانات.

قرار	النقاط الكمية لـ GMAT (بالنسبة المئوية)	GPA	GMAT
لا	٣٥	٢,٧٥	٦٥٠
لا	٧٠	٣,٥٠	٥٨٠
نعم	٧٥	٣,٥٠	٦٠٠
لا	٨٠	٢,٩٥	٤٥٠
نعم	٩٠	٣,٢٥	٧٠٠
نعم	٨٠	٣,٥٠	٥٩٠
لا	٤٥	٣,٨٥	٤٠٠
نعم	٧٥	٣,٥٠	٦٤٠

قرار	النقاط الكمية لـ GMAT (بالنسبة المئوية)	GPA	GMAT
؟	٦٠	٣,٠٠	٥٤٠
؟	٨٠	٢,٨٥	٦٩٠
؟	٦٥	٤,٠٠	٤٩٠

تمارين الإنترنت:

١- قــم بزيــارة AI Exploratorium في cs.ualberta.ca/aixplore. انقــر عــلى رابط شــجرة القــرارات. اقرأ السرد على إحصائيات لعبة كرة السَّــلة. افحص البيانات، ثم أنشــئ شــجرة قرار. اكتب تقريرًا عن انطباعاتك عن دقة شــجرة القرارات هذه. استكشــف أيضًا تأثيرات الخوارزميات المختلفة.

٢- قم باستقصاء بعض أدوات التنقيب في البيانات وموردِّيها. ابدأ بكلٍّ من fico.com وegain.com. قــم بزيارة موقــع dmreview.com، وحَدِّد بعــض منتجات التنقيــب في البيانات ومقدِّمي الخدمات غير المذكورة في هذا الفصل.

٣- ابحث عن حالات حديثة من تطبيقات التنقيب في البيانات الناجحة. قم بزيارة مواقع الويب لبعض موردِّي أدوات التنقيب في البيانات، وابحث عن حالات أو قصص نجاح. ثم قم بإعداد تقرير يُلخِّص خمس دراسات حالة جديدة.

٤- انتقــل إلى مواقع الويــب الخاصة بالبائعين (خاصــةً SAS، وSPSS، وCognos، وTeradata، وStatSoft، وFair Isaac) واطلــع عــلى قصــص النجــاح الخاصــة بــأدوات ذكاء الأعــمال (OLAP والتنقيب في البيانات). ما الذي تشترك فيه قصص النجاح المختلفة؟ كيف تختلف؟

٥- انتقــل إلى statsoft.com (وهــي إحدى شركات Dell الآن). قــم بتنزيل ثلاث ورقات بيضاء على الأقل حول التطبيقات. أيٌّ من هذه التطبيقات ربما اســتخدم تقنيات تعدين البيانات / النصوص / الويب التي تمَّت مناقشتها في هذا الفصل؟

٦- اذهب إلى sas.com، وقم بتنزيل ثلاث وثائق بيضاء على الأقل حول التطبيقات. أيٌّ من هذه التطبيقات ربما استخدم تقنيات تعدين البيانات / النصوص / الويب التي تمَّت مناقشتها في هذا الفصل؟

٧- انتقل إلى spss.com (شركة IBM). قم بتنزيل ثلاث ورقات بيضاء على الأقل حول التطبيقات. أيٌّ من هذه التطبيقات ربما اســتخدم تقنيات تعدين البيانات / النصوص / الويب التي تمَّت مناقشتها في هذا الفصل؟

٨- انتقل إلى teradata.com. قم بتنزيل ثلاث ورقات بيضاء على الأقل حول التطبيقات. أيٌّ من هذه التطبيقات ربما استخدم تقنيات تعدين البيانات / النصوص / الويب التي تمَّت مناقشتها في هذا الفصل؟

٩- انتقل إلى fico.com. قم بتنزيل ثلاث ورقات بيضاء على الأقل حول التطبيقات. أيٌّ من هذه التطبيقات ربما اســتخدم تقنيات تعدين البيانات / النصوص / الويب التي تمَّت مناقشــتها في هذا الفصل؟

١٠- انتقل إلى salfordsystems.com. قم بتنزيل ثلاث ورقات بيضاء على الأقل حول التطبيقات. أيٌّ من هذه التطبيقات ربما اســتخدم تقنيات تعدين البيانات / النصوص / الويب التي تمَّت مناقشتها في هذا الفصل؟

١١- انتقــل إلى rulequest.com. قــم بتنزيــل ثلاث ورقات بيضاء على الأقــل حول التطبيقات. أيٌّ من هذه التطبيقات ربما اســتخدم تقنيات تعدين البيانات / النصوص / الويب التي تمَّت مناقشتها في هذا الفصل؟

١٢- انتقل إلى kdnuggets.com. استكشف الأقسام على التطبيقات وكذلك البرامج. ابحث عن أسماء ثلاث حزم إضافية على الأقل للتنقيب في البيانات وتنقيح النصوص.

المراجع:

- Abbott, D. (2014). Applied predictive analytics: Principles and techniques for the professional data analyst. John Wiley & Sons.
- Anthes, G. H. (1999). "The next decade: interview with Arno A. Penzias," Computerworld, 33(1), pp. 3-4.
- Chan, P. K., Phan, W., Prodromidis, A., & Stolfo, S. (1999). Distributed data mining in credit card fraud detection. IEEE Intelligent Systems, 14(6), 67-74.
- CRISP-DM. (2013). Cross-Industry Standard Process for Data Mining (CRISP-DM). http://crisp-dm.orgwww.the-modeling-agency.com/crisp-dm.pdf (accessed February 2, 2013).
- Davenport, T. H. (2006, January). Competing on analytics. Harvard Business Review, 99-107.
- Delen, D. (2009). Analysis of cancer data: A data mining approach. Expert Systems, 26(1), 100-112.
- Delen, D. (2014). Real-world data mining: Applied business analytics and decision making. Upper Saddle River, NJ: Pearson.
- Delen, D., Cogdell, D., & Kasap, N. (2012). A comparative analysis of data mining methods in predicting NCAA Bowl outcomes. International Journal of Forecasting, 28, 543-552.
- Delen, D., & Sharda, R. (2010). Predicting the financial success of Hollywood movies using an information fusion approach. Industrial Engineering Journal, 21(1), 30-37. Delen, D., Sharda, R., & Kumar, P. (2007). Movie forecast Guru: A Web-based DSS for Hollywood managers. Decision Support Systems, 43(4), 1151-1170.
- Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. Artificial Intelligence in Medicine, 34(2), 113-127.
- Dunham, M. (2003). Data mining: Introductory and advanced topics. Upper Saddle River, NJ: Prentice Hall.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From knowledge discovery in databases. AI Magazine, 17(3), 37-54.
- Hoffman, T. (1998, December 7). Banks turn to IT to reclaim most profitable customers. Computerworld.
- Hoffman, T. (1999, April 19). Insurers mine for age-appropriate offering. Computerworld.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological Cybernetics, 43(1), 59-69.
- Nemati, H. R., & Barko, C. D. (2001). Issues in organizational data mining: A survey of current practices. Journal of Data Warehousing, 6(1), 25-36.
- Nisbet, R., Miner, G., & Elder IV, J. (2009). "Top 10 Data Mining Mistakes" in the Handbook of statistical analysis and data mining applications. pp. 733-754. Academic Press.
- Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1, 81-106.
- SEMMA. (2009). SAS's data mining process: Sample, explore, modify, model, assess.sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html (accessed August 2009).

- Seni, G., & Elder, J. F. (2010). Ensemble methods in data mining: Improving accuracy through combining predictions. Synthesis Lectures on Data Mining and Knowledge Discovery, 2(1), 1-126.
- Sharda, R., & Delen, D. (2006). Predicting box-office success of motion pictures with neural networks. Expert Systems with Applications, 30, 243-254.
- Shultz, R. (2004, December 7). Live from NCDM: Tales of database buffoonery. directmag.com/news/ncdm-12-07- 04/index.html (accessed April 2009).
- Skalak, D. (2001). Data mining blunders exposed! DB2 Magazine, 6(2), 10-13.
- Thongkam, J., Xu, G., Zhang, Y., & Huang, F. (2009). Toward breast cancer survivability prediction models through improving training space. Expert Systems with Applications, 36(10), 12200-12209.
- Wald, M. L. (2004, February 21). U.S. calls release of JetBlue data improper. The New York Times.
- Wright, C. (2012). Statistical predictors of March Madness: An examination of the NCAA Men's Basketball Championship. http://economics-files.pomona.edu/GarySmith/Econ190/Wright%20March%20Madness%20Final%20Paper.pdf (accessed February 2, 2013).
- Zaima, A. (2003). The five myths of data mining. What Works: Best practices in business intelligence and data warehousing, Vol. 15. Chatsworth, CA: The Data Warehousing Institute, pp. 42-43.
- Zolbanin, H. M., Delen, D., & Zadeh, A. H. (2015). Predicting overall survivability in comorbidity of cancers: A data mining approach. Decision Support Systems, 74, 150-161.

الفصل الخامس
التحليلات التنبؤية (٢)
تحليلات النص والويب ووسائل التواصل الاجتماعي

أهداف التعلم:

- وصف تحليلات النصوص وفهم الحاجة إلى التنقيب في النص.
- التمييز بين تحليلات النصوص، والتنقيب في النص، والتنقيب في البيانات.
- فهم المجالات المتنوعة لتطبيق التنقيب في النص.
- التعرف على عملية تنفيذ مشروع التنقيب في النص.
- التعرف على الطرق المختلفة لهيكلة البيانات النصية.
- وصف تحليل المشاعر.
- رفع مستوى الإلمام بالتطبيقات الشائعة لتحليل المشاعر.
- تعلم الطرق الشائعة لتحليل المشاعر.
- التمرس على تحليلات الكلام كما ترتبط بتحليل المشاعر.

يقــدم هذا الفصل نظرةً شــاملة على تحليلات / التنقيب في النصــوص / الويب، بالإضافة إلى مجالات التطبيق الشــائعة مثل محركات البحث وتحليل المشــاعر وتحليلات الشبكات / وسائل التواصل الاجتماعي. وكما شهدنا في السنوات الأخيرة، فإن البيانات غير المهيكلة التي يتم إنشاؤها عــبر الإنترنت (الويب، وشــبكات الاستشــعار، وأنظمة التعــرف على تــرددات الراديو [RFID]، وشــبكات المراقبة، إلخ) تزداد بسرعة هائلة وليس هناك مؤشر على تباطئها. تُجبِر هذه الطبيعة المتغيرة للبيانات المنظمات على جعل تحليلات النص والويب جزءًا حساسًا من بنيتها التحتية فيما يخص ذكاء / تحليلات الأعمال.

٥-١ مقال افتتاحي: الآلة مقابل الإنسان على Jeopardy: قصة واتسون:

هل تستطيع الآلة التغلب على أفضل ما في الإنسان فيما يفترض أن يكون الإنسان هو الأفضل فيه؟ يبدو أن الإجابة هي: نعم، واسم الآلة هو واتسون. حيث يعتبر واتسون هو نظام حاسب

استثنائي (مزيج جديد من الأجهزة والبرامج المتقدمة) مصمم للإجابة على الأسئلة المطروحة باللغة البشرية الطبيعية. وقد تم تطويره في عام ٢٠١٠ من قِبل فريق IBM Research كجزء من مشروع DeepQA واستمد اسمه من أول رئيس لشركة IBM، وهو Thomas J. Watson.

خلفية:

منذ ما يقرب من ٣ سنوات، كانت IBM Research تبحث عن تحدٍ بحثيٍّ كبير لمجاراة الاهتمامات العلمية والشعبية لـ Deep Blue، بطل لعبة الشطرنج على الحاسب، والذي قد يرتبط بعلاقة واضحة بالاهتمامات التجارية الخاصة بشركة IBM. كان الهدف هو تطوير علوم الحاسب من خلال استكشاف طرق جديدة لتقنية الحاسب تؤثر على العلوم والأعمال التجارية والمجتمع. وبناءً على ذلك، قبلت IBM Research تحديًا ببناء نظام حاسب يمكن أن ينافس على لقب بطل العالم بشكل فوري لبرنامج مسابقة التليفزيون الأمريكي Jeopardy! وقد شمل نطاق التحدي مشاركة متسابق آلي بشكل فوري للعرض التليفزيوني، ويكون لديه القدرة على الاستماع والفهم والاستجابة، وليس مجرد تمرين معملي.

التنافس ضد الأفضل:

في عام ٢٠١١، وكاختبار لقدراته، تنافس واتسون على مسابقة Jeopardy، في أول مباراة بين البشر والآلة. في مباراة ثنائية مجمعة النقاط (تم بثها عبر ثلاث حلقات من العرض التلفازي Jeopardy في الفترة ١٤-١٦ فبراير)، فاز واتسون على Brad Rutter، أكبر رابح بالمال على الإطلاق في Jeopardy، وعلى Ken Jennings، حامل الرقم القياسي لأطول فترة بالبطولة وهي (٧٥ يومًا). في هذه الحلقات، استمر واتسون في التفوق على خصومه من البشر وفقًا لجهاز إشارات اللعبة، لكنه واجه صعوبة في الرد على بعض الفئات، خاصةً أولئك الذين لديهم أدلة قصيرة تحتوي فقط على بضع كلمات. استطاع واتسون الوصول إلى ٢٠٠ مليون صفحة منظمة / غير منظمة المحتوى ومخزنة على ٤ تيرابايت من إجمالي السعة التخزينية للقرص. ولم يكن واتسون متصلًا بالإنترنت خلال المباراة.

أعلن أحد المختصين في لقاء تلفازي عن Jeopardy أن التحدي المطلوب هو التقدم ودمج مجموعة متنوعة من تقنيات ضمان الجودة (التنقيب في النصوص ومعالجة اللغات الطبيعية)، بما في ذلك التحليل اللفظي، وتصنيف الأسئلة وتحليلها واقتناء المصادر بصورة تلقائية وتقييمها، واكتشاف الكيان والعلاقة، وتكوين النماذج المنطقية، وتمثيل المعرفة والتفكير. وأضاف أن الفوز في Jeopardy يتطلب أن تتحلى إجاباتك بثقة في الحوسبة بصورة عالية الدقة. وفي الوقت الذي تكتسي فيه الأسئلة والمحتوى بالغموض والضجيج، ولا يتمتع أي من الخوارزميات الفردية بالكمال.

لذلـك؛ يجب أن يقدم كل مكون مخرجاته بدرجة من الثقة، كما يجب تجميع درجات الثقة لكل المكونـات الفردية للقيام بحسـاب الثقة الكلية للإجابـة النهائية. ويتم اسـتخدام الثقة النهائية لتحديد ما إذا كان ينبغي على نظام الحاسب المخاطرة باختيار الإجابة على أي حال. في لغة برنامج Jeopardy، تستخدم هذه الثقة لتحديد ما إذا كان الحاسب سوف يصدر "رنينًا" أو "طنينًا" لسؤال ما. وهنا يجب أن يتم حساب الثقة خلال الوقت الذي تتم فيه قراءة السؤال وقبل فرصة صدور الرنين للسؤال. ويقدر هذا الوقت تقريبًا ما بين ثانية واحدة إلى ٦ ثوانٍ بمتوسط ٣ ثوانٍ تقريبًا.

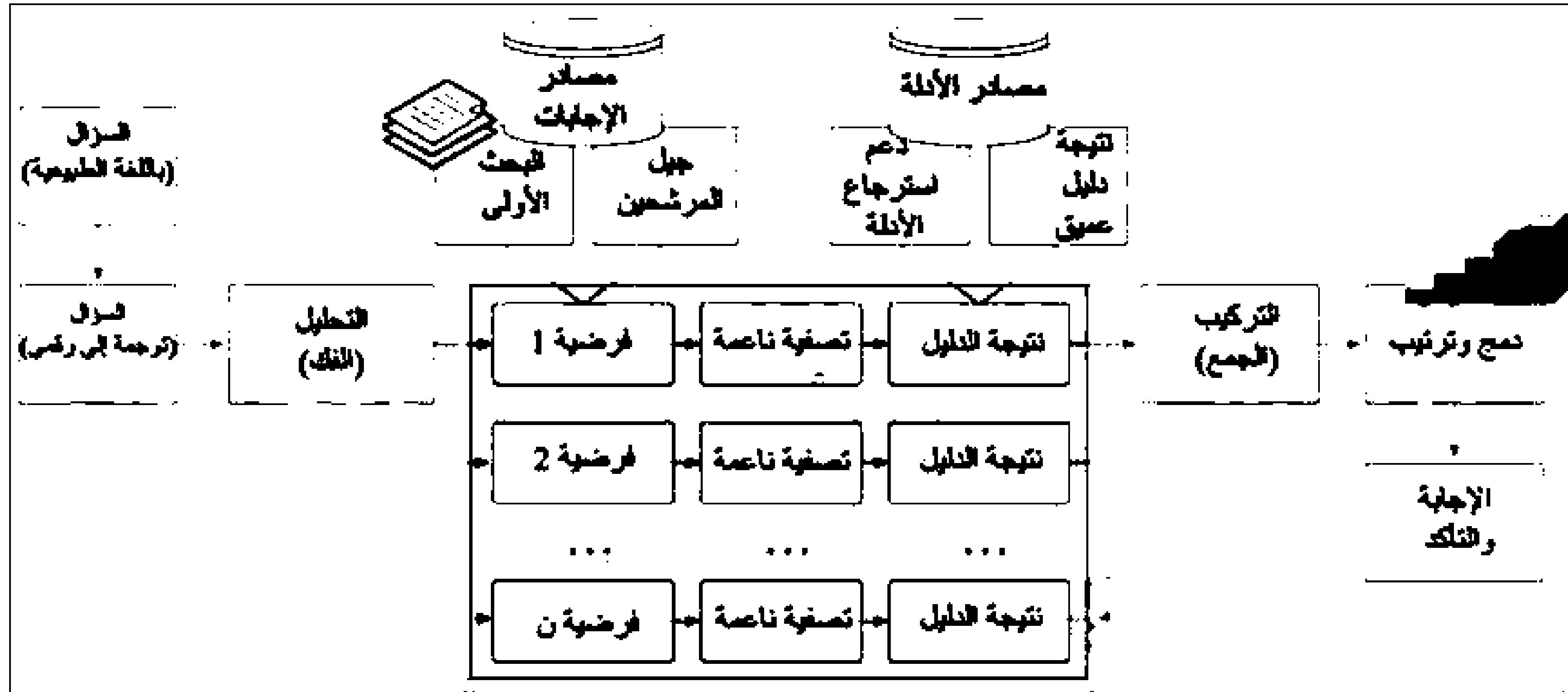
كيف يفعل واتسون ذلك؟

إن نظام واتســون، والذي يدعى DeepQA، هو بنية حســابية متوازية وواسعة النطاق تركز على التنقيب في النصوص، وتســتند إلى الأدلة الاحتمالية. وبالنســبة لتحدي Jeopardy، استخدم واتســون أكثر من ١٠٠ طريقة مختلفة لتحليل اللغــة الطبيعية، وتحديد المصادر، وإيجاد وتوليد فرضيــات، وإيجــاد الأدلة وإعطائها نقاط، ودمج وترتيب الفرضيــات. وما هو أهم بكثير من أي تقنية تم استخدامها هو كيفية دمجهم في DeepQA بحيث يمكن للطرق المتداخلة أن تستحضر نقاط قوّتها للإنتاج والمساهمة في تحسين كلٍ من الدقة والثقة والسرعة.

تعتبر DeepQA بنية ذات منهجية مصاحبة، وهي ليست خاصة بتحدي Jeopardy. وتشكل المبادئ الشــاملة في DeepQA توازيًا هائلًا، وعديدًا من الخبراء، وتقديرًا للثقة المنتشرة، وتكاملًا في تحليلات النص بصورة أحدث وأعظم.

- التوازي الهائل: يفيد التوازي الهائل في النظر في تفسيرات وفرضيات متعددة.
- العديد من الخبراء: يســهل التكامل والتطبيق والتقييم السياقي لمجموعة واسعة من تحليلات الأسئلة والمحتوى الاحتمالي المتباعدين عن بعضهما.
- تقدير الثقة المنتشرة: لا يلتزم أي عنصر بأي إجابة؛ فجميع المكونات تنتج ميزات ومستوى الثقة المرتبطة بها، وسجل تفسيرات مختلفة للأسئلة والمحتوى. حيث تقوم الركيزة الأساسية لمعالجة الثقة بالتعرف على كيفية تكديس النتائج ودمجها.
- تكامــل المعرفة الســطحية والعميقة: ويتم ذلك عن طريق الموازنة في اســتخدام الدلالات الصارمة والسطحية، والاستفادة من العديد من التجميعات التي تكونت بشكل يتسم بالأريحية.

يوضــح الشــكل (٥-١) بنية عالية المســتوى لــ DeepQA. ولمزيد مــن التفاصيل الفنية حول المكونات المختلفة لهذه البنية وأدوارها وقدراتها المحددة انظر: Ferrucci وآخرين (٢٠١٠).



شكل ٥-١: تصوير عالي المستوى لبنية DeepQA

استنتاج:

ساعد تحدي Jeopardy شركة IBM في جمع المتطلبات التي أدت إلى تصميم بنية DeepQA وتنفيذ برنامج واتسون. وبعد ٣ سنوات من البحث والتطوير المكثف من قِبل فريق أساسي يتكون تقريبًا من ٢٠ باحثًا، يعمل واتسون في العرض التلفازي لمسابقة Jeopardy على مستويات عالية من الخبرة البشرية من حيث الدقة والثقة والسرعة.

تزعم IBM أنها طورت العديد من الخوارزميات الحسابية واللغوية لمعالجة أنواع مختلفة من القضايا والمتطلبات فيما يخص ضمان الجودة. وعلى الرغم من أن بواطن هذه الخوارزميات غير معروفة، إلا أنها حققت أقصى استفادة من تحليلات النص والتنقيب في النص. وتعمل شركة IBM الآن على إصدار واتسون للتعامل مع المشكلات التي يمكن التغلب عليها في مجال الرعاية الصحية والطب (Feldman وآخرون، ٢٠١٢).

ماذا يمكننا أن نتعلم من هذه المقالة القصيرة:

مما لا شك فيه أن تقنية الحاسب - سواء الأجهزة أو واجهات البرامج - تتقدم بشكل أسرع من أي شيء آخر في الخمسين سنة الماضية. إن الأشياء التي كانت كبيرة جدًا ومعقدة جدًا ومن المستحيل حلها، أصبحت الآن في متناول تقنية المعلومات. وتعتبر تحليلات / التنقيب في النص إحدى التقنيات التمكينية المستخدمة في ذلك. وقد تم إنشاء قواعد بيانات لتركيب البيانات بحيث يمكن معالجتها بواسطة أجهزة الحاسب. من ناحية أخرى، كان المقصود بالنسبة للنص أن تتم معالجته بواسطة البشر. والسؤال الذي يطرح نفسه هنا، هل يمكن للآلات القيام بالأشياء التي تتطلب إبداع الإنسان وذكاءه، وهي لم تصمم أصلًا للآلات؟ من الواضح أن الإجابة: نعم.

فيعتبر واتسون مثالًا رائعًا على المسافة التي قطعناها في معالجة المستحيل. وقد أصبحت أجهزة الحاســب الآن ذكية بما يكفي للقيام بدور البشر فيما نعتقد أن البشر هم الأفضل في هذا المجال. إن فهم السؤال الذي تم طرحه في اللغة البشرية المنطوقة، ومعالجته وفهمه فهمًا جيدًا، والبحث عــن إجابة له، والرد عليه في غضون ثوانٍ قليلة كان شــيئًا لم يكــن بإمكاننا تصوره قبل أن يفعله واتســون فعلًا. وفي هذا الفصل، ستتعلم الأدوات والتقنيات التي قام عليها واتسون والعديد من الأجهزة الذكية الأخرى والتي صنعت المعجزات من أجل معالجة المشاكل التي كان يُعتقد في يوم من الأيام أنه من المستحيل حلها.

أسئلة مراجعة على المقالة الافتتاحية:

٦- ما هو واتسون؟ ما هو المميز فيه؟

٧- ما هي التقنيات المستخدمة في بناء واتسون (سواء من الأجهزة أو من البرامج)؟

٨ ما هي الخصائص المبتكرة لبنية DeepQA التي صنعت تفوق واتسون؟

٩- لماذا أنفقت IBM كل هذا الوقت والمال لبناء واتسون؟ أين عائد الاستثمار (ROI)؟

٥-٢ نظرة عامّة على تحليلات النص والتنقيب في النص:

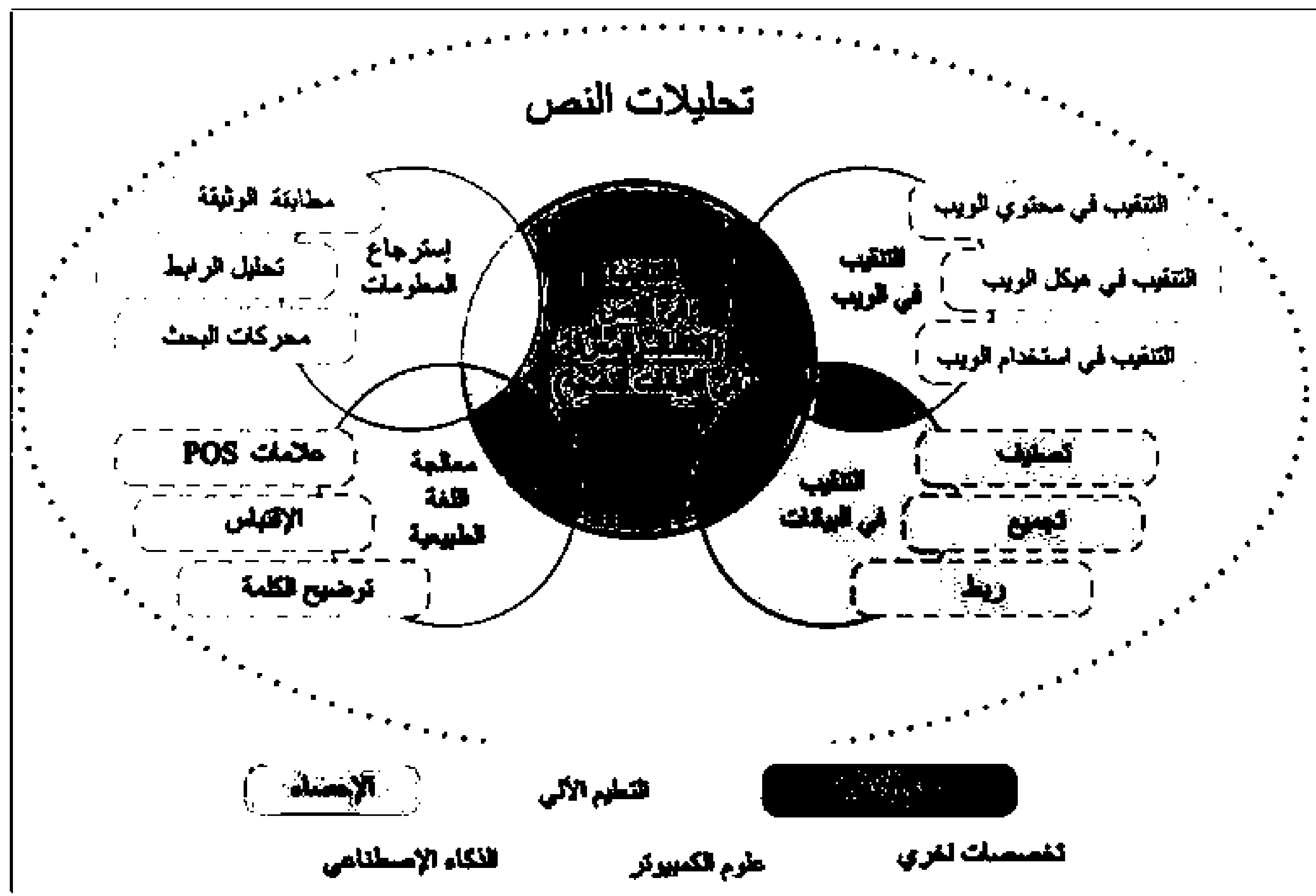
يتميــز عصر المعلومات الــذي نعيش فيه بالنمو السريع في كميــة البيانات والمعلومات التي يتم جمعها وتخزينها وإتاحتها في شــكل إلكتروني. حيث يتم تخزين الغالبية العظمى من بيانات الأعــمال في وثائــق نصية غير منظمــة فعليًا. ووفقًا لدراســة أجراها كل مــن Merrill Lynch وGartner، فــإن ٨٥٪ مــن جميع بيانــات الشركة يتــم التقاطها وتخزينها في شــكل غير منظم (McKnight, 2005). وقد ذكرت نفس الدراســة أيضًا أن هــذه البيانات غير المنظمة يتضاعف حجمها كل ١٨ شهرًا. ونظرًا لأن المعرفة تمثل قوة في عالم الأعمال اليوم، بالإضافة إلى أنها مستمدة من البيانات والمعلومات، فإن الشركات التي تســتفيد بكفاءة وبشــكل فعال من مصادر البيانات النصية يكون لديها المعرفة اللازمة لاتخاذ قرارات أفضل، مما يجعلها تتميز في المنافســة عن تلك الشركات التي تتخلف عن الركب. وهذا تماما هو المكان الذي تتناســب فيه الحاجة إلى تحليلات النصوص والتنقيب في النص مع الصورة الكبيرة الحالية للأعمال التجارية.

وعــلى الرغم من أن الهدف الرئيس لكل من تحليــلات النصوص والتنقيب في النص هو تحويل البيانات النصية غير المهيكلة إلى معلومات قابلة للتطبيق من خلال تطبيق معالجة اللغات الطبيعية (NLP) والتحليلات، حيث إن تعريفاتها مختلفة بعض الشيء، على الأقل بالنسبة لبعض خبراء هذا المجال. ووفقًا لهؤلاء الخبراء، فإن مفهوم تحليلات النص هو مفهوم أوســع حيث يشــمل استرجاع المعلومــات (مثل البحث وتحديد الوثائق ذات الصلة لمجموعة معينة من المصطلحات الرئيســة)،

وكذلك استخراج المعلومات، والتنقيب في البيانات، والتنقيب في الويب، في حين يركز التنقيب في النص في المقام الأول على اكتشاف معرفة جديدة ومفيدة من مصادر البيانات النصية. ويوضح الشكل (٥-٢) العلاقات بين تحليلات النصوص والتنقيب في النص إلى جانب مناطق التطبيق الأخرى ذات الصلة. ويبين الجزء السفلي من الشكل (٥-٢) التخصصات الرئيسة (أساس التصميم) التي تلعب دورًا حساسًا في تطوير مجالات التطبيق الأكثر انتشارًا بشكل متزايد. وبناءً على هذا التعريف لتحليل النصوص، والتنقيب في النص، فإنه يمكن ببساطة حساب الفرق بينهما كما يلي:

تحليلات النص = استرجاع المعلومات + استخراج المعلومات + التنقيب في البيانات + التنقيب في الويب.

أو ببساطة: تحليلات النص = استرجاع المعلومات + التنقيب في النص.



شكل ٥-٢: تحليلات النص والتطبيقات المرتبطة والتخصصات التمكينية

Sources: Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A.,... Welty, C. (2010). Building Watson: An overview of the DeepQA Project. AI Magazine, 31(3); DeepQA. DeepQA Project: FAQ, IBM Corporation (2011). research.ibm.com/deepqa/faq.shtml (accessed January 2013); Feldman, S., Hanover, J., Burghard, C., & Schubmehl, D. (2012). Unlocking the power of unstructured data. IBM white paper. www-01.ibm.com/software/ebusiness/jstart/downloads/unlockingUnstructuredData.pdf (accessed February 2013).

ويُعـد مصطلح تحليلات النصوص مصطلحًا جديدًا نسـبيًا بالمقارنـة مع مصطلح التنقيب في النـص؛ إذ يركز مصطلح تحليلات النصوص على التحليلات، وكما هو الحال في العديد من مجالات تطبيقـات التقنية الأخرى ذات الصلة (مثل تحليلات المسـتهلك، والتحليلات الكاملة، والتحليلات المرئية، والتحليلات الاجتماعية)، فإن مجال النص يحتاج أيضًا إلى الحصول على مُحاكاة للتحليلات. وعلى الرغم من استخدام مصطلح تحليلات النصوص بشكل أكثر انتشارًا في سياق تطبيق الأعمال، إلا أن مصطلح التنقيب في النص يسـتخدم بشـكل متكرر في دوائر البحث الأكاديمية. ورغم أنه في بعـض الأحيان قد يتم تعريف مصطلحي (تحليلات النص والتنقيب في النص) بطريقة مختلفة بعض الشيء، إلا أنهما يستخدمان بشكل مترادف.

التنقيب في النص (ويُعرف أيضًا باسـم التنقيب في البيانات النصية، أو اكتشاف المعرفة في قواعد البيانات النصية) هو عملية شـبه آلية لاسـتخلاص الأنماط (المعلومات والمعرفة المفيدة) من كميات كبـيرة من مصادر البيانات غير المُهيكلـة. والجدير بالذكر أن التنقيب في البيانات هو عملية تحديد أنمـاط صالحة ومبتكرة ومفيـدة ومفهومة في بيانات مخزنة في قواعـد البيانات المنظمة، حيث يتم تنظيم البيانات في سجلات منظمة من خلال متغيرات فئوية أو ترتيبية أو مستمرة. ويُعد التنقيب في النص هو نفسه التنقيب في البيانات من حيث استهدافه لنفس الغرض واستخدامه لنفس العمليات، ولكـن مع التنقيـب في النص ينطوي الإدخـال إلى العملية على مجموعة من ملفـات البيانات غير المهيكلة (أو الأقل تنظيمًا)، مثل: وثائق Word، وملفات PDF والمقتبسـات النصية وملفات XML، وما إلى ذلك. ويمكن اعتبار التنقيب في النص في جوهره كعملية (ذات خطوتين رئيسيتين) تبدأ بفرض هيكلة مصادر البيانات المسـتندة إلى النص يتبعها اسـتخلاص المعلومـات والمعارف ذات الصلة من هذه البيانات المهيكلة والمستندة إلى النص، وذلك باستخدام تقنيات وأدوات التنقيب في البيانات.

وتظهـر فوائد التنقيب في النـص بوضوح في المجالات التي يتم فيهـا توليد كميات كبيرة من البيانات النصية، مثل القانون (أوامر المحكمة)، والبحوث الأكاديمية (المقالات البحثية)، والتمويل (تقاريـر فصلية)، بالإضافة إلى الطـب (ملخصات التفريغ)، وعلم الأحيـاء (التفاعلات الجزئية)، والتقنية (ملفات براءات الاختراع)، فضلاً عن التسـويق (تعليقات العملاء). فعلى سـبيل المثال، يمكن اسـتخدام التفاعلات النصية القائمة على النموذج الحر مع العملاء في شـكل شـكاوى (أو مدح) ومطالبات الضمان من أجل تحديد خصائص المنتجات والخدمات التي تعتبر أقل من مثالية بشـكل موضوعي، ويمكن اسـتخدامها كمدخلات لتطوير منتجات أفضـل ومخصصات الخدمة. وبالمثل، فإن برامج التوعية السـوقية ومجموعات التركيز تولد كميات كبيرة من البيانات. ويمكن للعملاء تقديم أفكارهم حول منتجات وخدمات الشركة بأسـلوبهم الخاص، وذلك من خلال عدم

تقييـد التعليقات على المنتجـات أو الخدمات أو جعلها مقننة. ومن المجالات الأخرى التي أثرت فيهـا المعالجة الآلية للنص غير المنظم تأثيرًا كبـيرًا مجال الاتصالات الإلكترونية والبريد الإلكتروني. والجديـر بالذكر أن التنقيب في النص لا يُسـتخدم فقط لتصنيف وتصفيـة البريد الإلكتروني غير الهام، وإنما يمكن اسـتخدامه أيضًا لتحديد أولويات البريد الإلكتروني تلقائيًا اسـتنادًا إلى مسـتوى الأهميـة بالإضافة إلى إنشـاء ردود تلقائيـة (Weng & Liu, 2004). ونذكـر فيما يلي مجالات التطبيق الأكثر شيوعًا في التنقيب في النصوص:

- استـخراج المعلومات: تحديد العبارات والعلاقات الرئيسـة داخل النص من خلال البحث عن الكائنات والتسلسلات المحددة مسبقًا في النص عن طريق مطابقة النمط.
- تتبـع الموضوع: اسـتنادًا إلى ملف تعريف المُسـتَخدِم بالإضافـة إلى الوثائق التي تحتوي على وجهات نظر المُستَخدِم، فإن التنقيب في النص يمكنه التنبؤ بالوثائق الأخرى التي تهم المستخدم.
- التلخيص: تلخيص الوثيقة توفيرًا للوقت من جانب القارئ.
- التصنيف: تحديد الموضوعات الرئيسة للوثيقة ثم وضعها في مجموعة كبيرة محددة مسبقًا من الفئات بناءً على تلك الموضوعات.
- التجميع: تجميع وثائق مشابهة بدون وجود مجموعة محددة مسبقًا من الفئات.
- ربط مفهوم: ربط الوثائق ذات الصلة من خلال تحديد المفاهيم المشتركة، وبذلك يساعد المستخدمين في الوصول إلى المعلومات التي قد يصعب الوصول إليها باستخدام الطرق التقليدية.
- إجابة السؤال: إيجاد أفضل إجابة لسؤال معين من خلال مطابقة النمط القائم على المعرفة.

وللحصول على شرح لبعض المصطلحات والمفاهيم المُستخدَمة في التنقيب في النص انظر رؤية فنية (٥-١)، في حين توضح الحالة العملية (٥-١) استخدام التنقيب في النص في صناعة التأمين.

رؤية فنية ٥-١

مصطلحات التنقيب في النص

توضح القائمة التالية بعض مصطلحات التنقيب في النص شائعة الاستخدام

- بيانـات غير مهيكلـة (مقابل البيانات المهيكلة): للبيانات المهيكلة صيغة محددة مسـبقًا، وعـادةً ما يتم تنظيمها في سـجلات ذات قيم بيانات بسـيطة (فئويـة وترتيبية ومتغيرات مستمرة) ومخزنة في قواعد البيانات. وفي المقابل، فإن البيانات غير المهيكلة لا تحتوي على صيغة محددة مسـبقاً ويتم تخزينها في شـكل وثائق نصية. وتسـتخدم أجهزة الحاسوب

البيانـات المهيكلـة في جوهرها للقيام بالمعالجة، في حين تُسـتخدم البيانـات غير المهيكلة بواسطة البشر للمعالجة والفهم.

- مجموعـة القوانين Corpus: (جمـع corpora) وتُعرف لغويًا بأنها مجموعة كبيرة ومنسـقة من النصوص (التي عادةً ما يتم تخزينها ومعالجتها إلكترونيًا) تم إعدادها بغرض إجراء اكتشاف المعرفة.
- المصطلحات: المصطلح هو عبارة عن كلمة واحدة أو عبارة متعددة الكلمات مستخرجة مباشرة من مجموعة قوانين خاصة بمجال معين عن طريق أساليب معالجة اللغات الطبيعية NLP.
- المفاهيـم: هي مواصفات يتم إنشـاؤها مـن مجموعة من الوثائق عن طريق الوسـائل اليدويـة أو منهج تصنيف إحصائي أو قائم على القواعد أو مزيج ما بين كل ذلك. ومقارنة بالمصطلحات، تكون المفاهيم نتيجة لفكرة تجريدية على مستوى أعلى.
- الاشـتقاق Stemming: وهو عملية تقليـل الكلمات المصرفة إلى أصلها (أو قاعدتها أو جذرها). على سـبيل المثال: stemmer, stemming, stemmed كلها كلمات مشـتقة من الأصل وهو stem.
- كلـمات التوقف: كلمات التوقف (أو كلمات الضجيـج) وهي كلمات يتم تصفيتها قبل أو بعـد معالجـة بيانات اللغة الطبيعيـة (بمعنى النص). وعلى الرغـم من عدم وجود قائمـة مقبولة عالميًا لكلمات التوقف، فإن معظم أدوات معالجة اللغات الطبيعية NLP تستخدم قائمة تتضمن (a, am, the, of، إلخ)، والأفعال المساعدة (is, are, was, were، إلخ)، والكلمات المتعلقة بالسياق والتي تعتبر غير مميزة.
- المرادفات والمتجانسـات: المرادفات هي كلـمات مختلفة نحويًا (أي مختلفة هجائيًا) لكن ذات معانٍ متماثلة أو على الأقل متشـابهة (مثل السـينما والأفلام والصور المتحركة). وفي المقابل، فإن المتجانسات، والتي تسمى أيضًا المشتركات اللفظية، هي كلمات متطابقة نحويًا (أي متشابهة تمامًا من الناحية الهجائية) لكنها ذات معانٍ مختلفة (فمثلًا، القوس يمكن أن يعني "الانحناء إلى الأمام"، أو "واجهة السفينة"، أو "السلاح الذي يطلق السهام"، أو "نوع من الأشرطة المربوطة").
- الترميـز Tokenizing: الرمـز هو عبـارة نصية يتم تصنيفها في جملـة وفقًا للوظيفة التي تؤديهـا. ويعرف هذا التصنيـف بالترميز tokenizing. قد يبدو الرمز مثل أي شيء يحتاج أن يكون جزءًا نافعًا من النص المهيكل.
- قامـوس المصطلحـات: وهو مجموعة من المصطلحات الخاصـة بمجال محدد والتي يمكن استخدامها لحصر المصطلحات المستخرجة من داخل مجموعة القوانين (corpus).

- تكرار الكلمات: وهو عدد مرات العثور على كلمة ما في وثيقة ما.
- وضـع علامـات لجزء من الكلام Part-of-speech tagging: وهو عملية وسـم كلمات في النص كمطابقة لجزء معين من الكلام (مثل الأسماء والأفعال والصفات وظرف الحال والزمان والمكان، وما إلى ذلك) بناءً على تعريف الكلمة والسياق الذي تستخدم فيه.
- المورفولوجيا Morphology: وهو فرع من علوم اللغة وجزء من معالجة اللغات الطبيعية (NLP) والذي يدرس الهيكل الداخلي للكلمات (أنماط تنسيق الكلمة داخل لغة ما أو فيما بين عدة لغات).
- مصفوفـة Term-by-document (مصفوفة الأحداث): وهي مخطط تمثيلي شـائع للعلاقة القائمة على التكرار بين المصطلحات والوثائق في شـكل جدول حيث يتم وضع المصطلحات في أعمدة، والوثائق في صفوف، ويكون التكرار بين المصطلحات والوثائق في الخلايا كقيم صحيحة.
- فك القيمة المفرد (فهرسـة دلالات الألفاظ غير المباشرة): وهي طريقة تخفيض الأبعاد وتسـتخدم لتحويل مصفوفة term-by-document إلى حجم مقبول عن طريق توليد تمثيل متوسط للتكرارت باستخدام طريقة معالجة بارعة للمصفوفة وهي طريقة مشابهة لتحليل المكونات الأساسية.

حالة عملية ٥-١

مجموعة التأمين تعمل على تقوية إدارة المخاطر مع حل التنقيب في النص

عندما تم طرح سـؤال عن التحدي الأكبر الذي يواجه صناعة التأمين على السيارات التشـيكية، لم يتردد دكتور Peter Jedlicˇka رئيس فريـق الخدمات الاكتوارية لمكتب التأمينات التشيكية (CIB) في التعبير عن ذلك بقوله: «تزداد مطالبات الإصابة الجسدية بشكل غير متناسب مقارنةً بمطالبات تلف السيارة". وCIB هي منظمة مهنية لشركات التأمين في جمهورية التشيك والتي تتعامل مع المطالبات غير المؤمن عليها والدولية وغير المدونة لما يعرف باسـم مسؤولية الطرف الثالث للسيارات. ويضيف Peter Jedlicˇka "وتمثل تعويضات الإصابة الجسدية الآن حوالي ٤٥٪ من المطالبات المقدمة ضد أعضائنا، وستستمر هذه النسبة في التزايد بسبب التغييرات التشريعية الأخيرة".

وتتمثل إحـدى الصعوبات التـي تطرحها مطالبات الإصابة الجسـدية على شركات التأمـين في صعوبة التنبؤ بحجم الضرر المباشر من جراء حادث سـيارة؛ إذ تتفاقم بعض الإصابـات مع الوقت والتي لم تكن بتلك الحدة وقت الحادث، كما أن الإصابات التي قد تبدو طفيفة يمكن أن تتحول إلى حالات مزمنة. وفي السـابق كانت شركات التأمين تقدِّر بدقة مسؤوليتها عن الأضرار الطبية، وبصورة أدق كانت تستطيع إدارة مخاطرها وتوحيد مواردهـا. ومع ذلك، فإنه نظرًا لأن المعلومـات المطلوبة موجودة في وثائق غير مهيكلة

مثل تقارير الحوادث وتصريحات الشهود، فإن الأمر يستغرق وقتًا طويلًا للغاية بالنسبة للموظفين الأفراد لإجراء التحليل اللازم.

ولتوسيع وإجراء تحليل آلي لكلٍ من تقارير الحوادث غير المهيكلة، وعبارات الشهود، وصيــغ المطالبــات، قامت CIB بنشر حل تحليل البيانات اســتنادًا عــلى كل من Dell Statistica Data Miner وامتــداد Statistica Text Miner. حيــث تقدم Statistica Data Miner مجموعة أدوات بديهية وســهلة الاســتخدام، بحيث يمكن الوصول إليها حتى من قِبَل غير المحللين.

ويعمل الحل على قراءة وكتابة البيانات بشكل عملي من جميع التنسيقات المعيارية للملف كما يعرض أدوات تنظيف بيانات قوية ومتطورة. كما أنه يدعم حتى المستخدمين المبتدئين باســتخدام معالجات الاستعلام، والتي تُســمى وصفات التنقيب عن البيانات، والتي تساعدهم في الوصول إلى الإجابات التي يحتاجونها بشكل أسرع.

ومع امتداد Statistica Text Miner، يصل المستخدمون إلى أدوات الاستخراج والاختيار اللازمة لفهرسة وتصنيف وتجميع المعلومات من مجموعات كبيرة من البيانات النصية غير المهيكلة، مثل صيغ مطالبات التأمين. وبالإضافة إلى استخدام حل Statistica لعمل تنبؤات حول مطالبات الأضرار الطبية المســتقبلية، فيمكن استخدام CIB أيضًا لإيجاد الأنماط التي تكشف محاولات الاحتيال أو لتحديد التحسينات المطلوبة للسلامة على الطريق.

تحسينات في دقة تقديرات المسؤولية:

ويتوقع Jedlicˇka أن يقوم حل Statistica بتحســين قدرة CIB بشــكل كبير على التنبــؤ بالمطالبات الطبية الكلية التي قد تنشــأ عن أي حــادث؛ إذ يقول: "إن قدرات التنقيب في البيانات في مجال الإحصاء والتنقيب في النصوص تساعدنا بالفعل على كشف خصائــص مخاطر إضافية، مما يجعل من الممكن التنبــؤ بالادعاءات الطبية الخطيرة في المراحــل الأولى من التحقيق، ومع حل Statistica، يمكننا تقديم تقديرات أكثر دقة من إجمالي الأضرار والتخطيط وفقًا لذلك".

توسعات في عروض الخدمات للأعضاء:

ومــما زاد من ســعادة Jedlicˇka أيضًا من حــل Statistica هو أنه مكّن CIB من تقديم خدمات إضافية لأعضاء شركاتها، حيث يقول: "نحن في نشاط قائم على البيانات. ومع Statistica، يمكننا تزويد أعضائنا بتحليلات مفصلة للمطالبات واتجاهات السوق. كما تساعدنا Statistica في تقديم توصيات أقوى بشأن احتياطيات المطالبات".

البديهية بالنسبة لمستخدمي الأعمال:

يمكــن الوصول إلى أدوات Statistica البديهية حتى من قِبَل المســتخدمين غير الفنيين. وفي هذا الصدد يقول Jedlicˇka: "من السهل فهم مخرجات تحليلات الإحصائيات بالنسبة لمستخدمي الأعمال. ويكتشف مستخدمو الشركات أيضًا أن نتائج التحليل تسير على نفس نهج تجاربهم وتوصياتهم، لذا فبإمكانهم أن يشعروا بسهولة بقيمة حل Statistica".

أسئلة للمناقشة:

١- كيف يمكن اســتخدام تحليلات النص والتنقيب فيه لمواكبة الاحتياجات التجارية المتغيرة لشركات التأمين؟

٢- ما هي التحديات والحل المقترح والنتائج التي تم الحصول عليها؟

٣- هل يمكنك التفكير في اســتخدامات أخرى لتحليلات النصوص والتنقيب في النص لشركات التأمين؟

Sources: Dell Statistica Case Study. Insurance group strengthens risk management with text mining solution. https://software.dell.com/casestudy/czech-insurers-bureau-insurance-group-strengthens- risk-management-with875134/(accessed June 2016). Used by permission from Dell.

أسئلة مراجعة على القسم ٥-٢:

١- ما هي تحليلات النص؟ وكيف تختلف عن التنقيب في النص؟

٢- ما هو التنقيب في النص؟ وكيف يختلف عن التنقيب في البيانات؟

٣- لماذا يزداد انتشار التنقيب في النص كأداة تحليلية؟

٤- ما هي بعض مجالات التطبيق الأكثر انتشارًا في التنقيب في النص؟

٥-٣ معالجة اللغات الطبيعية (NLP):

تســتخدم بعض تطبيقات التنقيب المبكر في النص تمثيلًا مبسطًا يُسمى حقيبة الكلمات وذلك لإضفــاء الهيكلة على مجموعة من الوثائق النصية لتصنيفها إلى فئتين أو أكثر من الفئات المحددة مسبقًا أو تجميعها في تجمعات طبيعية. وفي نموذج حقيبة الكلمات، يتم تمثيل النص - جملة أو فقــرة أو وثيقــة كاملة - كمجموعة من الكلمات، دون تجاهــل قواعد أو ترتيب ظهور الكلمات.

ولايــزال نمــوذج حقيبة الكلمات يســتخدم في بعض أدوات تصنيف الوثائق البســيطة. فمثلًا في تصفية الرســائل غير المرغوب فيها، يمكن تصميم رسالة البريد الإلكتروني كمجموعة غير مرتبة من الكلمات (حقيبة كلمات) تتم مقارنتها بحقيبتين مختلفتين محددتين مســبقًا، حيث تمتلئ حقيبة واحدة بالكلمات الموجودة في رســائل البريد غــير المرغوب فيه والأخرى مليئة بكلمات تم العثور عليها في رسائل البريد الإلكتروني الاعتيادية. وعلى الرغم من أنه من المتوقع وجود بعض الكلمات في كلا الحقيبتــين، فإن حقيبة "البريد غير المرغوب فيه" تتكرر فيها كلمات تتعلق بالرســائل غير المرغــوب فيها - مثل المخزون، والفياجرا، والشراء - أكثر من تكرارها في الحقيبة الاعتيادية، والتي ســتحتوي على المزيد من الكلمات المتعلقة بأصدقاء المســتخدم أو مكان العمل. ويقوم مستوى التماثل بين حقيبة الكلمات الخاصة بالبريد الإلكتروني والحقيبتين اللتين تحتويان على التوصيفات بتحديد البريد الإلكتروني على أنها إما رسائل غير مرغوب فيها أو اعتيادية.

وبطبيعــة الحال، لا نســتخدم نحن (البشر) كلمات غير ذات ترتيب أو هيكل، بل نســتخدم كلــمات في جمل لها بنية دلالية ونحويــة. وبالتالي، فإنه يجب أن تبحــث التقنيات الآلية (مثل التنقيــب في النص) عن طرق لتجاوز تفســير حقيبة الكلمات وتضمين بنيــة دلالية أكثر وأكثر في عملياتهــا. ويتجه التنقيب في النص حاليًا نحو تضمــين العديد من المميزات المتقدمة التي يمكن الحصول عليها باستخدام معالجة اللغات الطبيعية NLP.

وقــد اتضح أن طريقة حقيبة الكلمات قد لا تنتج محتوى معلومات كافٍ بشــكل جيد لمهام التنقيــب في النص (مثــل التصنيف، والتجميع، والاقتران). وخير مثــال على هذا، يمكننا الحصول عليــه في الطــب القائم على الأدلة. ويتمثل أحد المكونات الحساســة للطب القائم على الأدلة في دمج أفضل نتائج البحوث المتاحة في عملية اتخاذ القرارات الإكلينيكية، والتي تنطوي على تقييم المعلومات التي تم جمعها من وســائل التواصل المطبوعة بالنسبة لصلاحيتها ومدى صلتها. وقد طور العديد من الباحثين من جامعة ميريلاند نماذج تقييم الأدلة باستخدام طريقة حقيبة الكلمات (Lin & Demner-Fushman, 2005). وقد استخدموا أساليب تعلم الآلة الشائعة بالإضافة إلى أكثر من نصف مليون بحث تم جمعها من MEDLINE (نظام تحليل وتحرير أدبيات الطب على الإنترنــت). وقد قاموا بتمثيل كل ملخص كحقائب كلــمات في النماذج التي قدموها، حيث كان كل مصطلح محدد يمثل ميزة. وعلى الرغم من استخدام أساليب التصنيف الشائعة مع منهجيات التصميــم التجريبي المجربة، فلم تعدو نتائج التنبؤ الخاصة بهم كونها مجرد تخمين بســيط، مما يشــير إلى أن حقيبــة الكلمات لا تنتج تمثيلًا جيدًا بما فيه الكفاية للمــواد البحثية في هذا المجال؛ وبالتالي فإن هناك حاجة إلى مزيد من التقنيات المتقدمة مثل معالجة اللغات الطبيعية NLP.

وتعتبر معالجة اللغات الطبيعية (NLP) أحد المكونات المهمة في التنقيب في النصوص، وهي عبارة عن مجال فرعي من الذكاء الاصطناعي وعلوم لغات الحاسوب. كما أنها تدرس مشكلة "فهم" لغة الإنسان الطبيعية، من وجهة نظر تحويل الصور البشرية (كالوثائق النصية) إلى تمثيلات أكثر رسمية (في شكل بيانات رقمية ورمزية) تسهل على برامج الحاسب إجراء المعالجات المطلوبة ببراعة. والهدف من معالجة اللغات الطبيعية هو إجراء المعالجة البارعة للنصوص المبنية على بناء الجملة (والذي غالبًا ما يطلق عليه "عد الكلمات") بشكل يوصل إلى فهم حقيقي ومعالجة اللغة الطبيعية التي تدرس القيود النحوية والدلالية وكذلك السياق.

ويعتبر تعريف ونطاق كلمة الفهم أحد موضوعات النقاش الرئيسية في معالجة اللغات الطبيعية NLP. وبالنظر إلى أن اللغة البشرية الطبيعية مبهمة، وأن الفهم الحقيقي للمعنى يتطلب معرفة واسعة بموضوع ما (يتجاوز ما هو موجود في الكلمات والجمل والفقرات)، فهل ستتمكن أجهزة الحاسب من فهم اللغة الطبيعية بنفس الطريقة وبنفس الدقة التي يفعلها البشر؟ والإجابة هي لا على الأغلب! لقد قطعت معالجة اللغات الطبيعية NLP شوطًا طويلًا منذ وقت استخدام العد البسيط للكلمات، غير أن هذه المعالجة للغات الطبيعية لديها طريقة أطول لتتمكن من الفهم الحقيقي للغة الإنسان الطبيعية. ونذكر فيما يلي عددًا قليلًا من التحديات المرتبطة بشكل عام بتنفيذ معالجة اللغات الطبيعية NLP:

- علامات جزء من الكلام: من الصعب ترميز المصطلحات في النص على أنها تطابق جزءًا معينًا من الكلام (مثل الأسماء والأفعال والصفات والظروف) لأن جزء الكلام لا يعتمد فقط على تعريف المصطلح ولكن أيضًا على السياق الذي يتم استخدامه فيه.
- تقسيم النص: بعض اللغات المكتوبة، مثل الصينية واليابانية والتايلاندية لا تحتوي على حدود مفردة للكلمات. وفي هذه الحالات فإن مهمة تحليل النص تتطلب تحديد حدود الكلمات، والتي غالبًا ما تكون مهمة صعبة، حيث تظهر تحديات مماثلة في تجزئة الكلام عند تحليل اللغة المحكية لأن الأصوات التي تمثل الحروف والكلمات المتعاقبة تمتزج مع بعضها البعض.
- توضيح معنى كلمة: حيث توجد كلمات كثيرة لها أكثر من معنى واحد، ولا يمكن أن يتحقق اختيار المعنى الأكثر اعتيادية إلا بمراعاة السياق الذي تستخدم فيه الكلمة.
- الغموض النحوي: حيث إن قواعد اللغة للغات الطبيعية غامضة؛ ولذلك فإن العديد من هياكل الجملة الممكنة في كثير من الأحيان تحتاج إلى النظر فيها، حيث إن اختيار البنية الأكثر ملاءمة يتطلب عادةً انصهار المعلومات الدلالية والسياقية.

- عـدم الكمال أو عدم انتظـام المدخلات: اللهجات الأجنبية أو الإقليميـة والعوائق الصوتية في الكلام والأخطاء المطبعية أو النحوية في النصوص تجعل معالجة اللغة مهمة أكثر صعوبة.
- أفعال لفظية: يمكن اعتبار الجملة في كثير من الأحيان إجراء من قِبَل المتحدث. فقد لا يحتوي هيكل الجملة وحده على معلومات كافية لتحديد هذا الإجراء. فعلى سبيل المثال: "هل يمكنك اجتياز الصف؟" يتطلب إجابة بسيطة؛ نعم / لا، في حين أن "هل يمكنك تمرير الملح؟" هو طلب لإجراء مادي يجب القيام به.

لقـد طال انتظار هذا الحلم من جانب مجتمع الذكاء الاصطناعي في إيجاد خوارزميات قادرة على القراءة والحصول على المعرفة من النص. ومن خلال تطبيق خوارزمية تعمل على نص معقد، قام باحثون من مختبر NLP بجامعة سـتانفورد بتطوير طرق يمكنها تحديد المفاهيم والعلاقات بين تلك المفاهيم في النص تلقائيًا. ومن خلال تطبيق إجراء فريد على كميات كبيرة من النصوص، فإن خوارزمياتها تحصل تلقائيًا على مئات الآلاف من عناصر المعرفة العالمية وتسـتخدمها لإنتاج مسـتودعات معززة بشـكل كبير لـ WordNet. وWordNet هي قاعدة بيانات مشـفرة يدويا مـن الكلمات الإنجليزية، وتعريفاتها، ومجموعات من المرادفات، والعلاقات الدلالية المختلفة بين المجموعات المترادفة، وهي مورد رئيسي لتطبيقات معالجة اللغات الطبيعية، ولكن ثبت أن تكلفة بنائهـا وصيانتها اليدوية مرتفعة للغاية. ومن خـلال إحداث المعرفة تلقائيًا في WordNet، فإنه توجـد إمكانية لجعل WordNet موردًا أكبر وأكثر شـمولًا لمعالجة اللغات الطبيعية بجزء صغير مـن التكلفة. وتُعد إدارة علاقـات العملاء (CRM) هي أحد المجالات البـارزة التي تظهر فيها فوائـد معالجة اللغات الطبيعية (NLP) وWordNet بالفعل. وبوجهٍ عام، فإن CRM يسـعى بجدية إلى تقدير العملاء من خلال فهم أفضل واسـتجابة فعالة لاحتياجاتهم الفعلية والمتوقعة. ويُعد تحليل المشاعر أحد المجالات المهمة في إدارة علاقات العملاء (CRM)، حيث يكون لمعالجة اللغـات الطبيعية (NLP) تأثير كبير في تحليل المشـاعر. وتحليل المشـاعر هو تقنية تسـتخدم لكشف الآراء المؤيدة وغير المؤيدة تجاه منتجات وخدمات معينة باستخدام عدد كبير من مصادر البيانات النصية (ملاحظات العملاء في شـكل منشـورات على شبكة الإنترنت). وسوف نتناول في القسم ٥-٦ تحليل المشاعر وWordNet بشيء من التفصيل.

ويمكن اسـتخدام التحليلات بشـكل عام وتحليلات النصوص والتنقيب في النص بشكل خاص في صناعة البث الإذاعي. وتقدم الحالة العملية ٥-٢ مثالًا على ذلك حيث يتم استخدام نطاق واسع من إمكانات التحليلات لالتقاط مشاهدين جدد وتوقع التقييمات وإضافة قيمة أعمال إلى شركات البث.

حالة عملية ٥-٢

استخدام شبكات AMC التحليلات لجذب مشاهدين جدد وتقييم التوقعات وإضافة القيمة للمعلنين في عالم متعدد القنوات

على مدى العشر سنوات الماضية كانت صناعة التلفاز المعتمد على كيبل في الولايات المتحدة تتمتع بفترة من النمو أتاحت إبداعًا غير مسبوق في إنشاء محتوى عالي الجودة. كانت شركة AMC Networks في طليعة هذا العصر الذهبي الجديد للتلفاز، مما أدى إلى إنتاج سلسلة من العروض المشهود لها بالنجاح بشكل حاسم مثل: Breaking Bad وMad Men وThe Walking Dead.

وتمتلــك شركــة AMC Networks المتخصصــة في إنتاج محتوى مــرئي من البرامج والأفلام بجودة عالية منذ أكثر من ٣٠ عامًا، العديد من العلامات التجارية الأكثر شعبية والحائــزة على جوائز في مجال التلفــاز المعتمد على كيبل، والتــي تنتج وتقدّم محتوى مميزًا وممتعًا وثقافيًّا يجذب الجماهير عبر منصات متعددة.

تصدر اللعبة:

وعلى الرغم من نجاح شركة AMC Networks، إلا أنها لا تمتلك خططًا للبقاء على أمجادها. وقد وضح Vitaly Tsivin، الباحث عن برنامج SVP Business Intelligence ذلــك بقولــه: "نحن لا نقف هكذا بــدون تقدم. فعلى الرغم من أن نســبة كبيرة من أعمالنا لا تزال تلفازًا خطيًّا، إلا أننا نحتاج إلى جذب جيل جديد من جيل الألفية الذين يستهلكون المحتوى بطرق مختلفة جدًا. لقد تطور التلفاز ليصبح عملًا متعدد القنوات ومتعدد الأنظمة، ومن ثَم فقد أصبحت الشــبكات الســلكية بحاجة إلى أن تكون أكثر ذكاءً حول كيفية تســويقها والتواصل مع الجمهور عبر جميع تلك التدفقات. والجدير بالذكر أن الاعتماد على بيانات التقديرات التقليدية ومقدمي تحليلات الجهات الخارجية تُعد إســتراتيجية خاسرة: تحتاج إلى الحصول على ملكية بياناتك، واستخدامها للحصول على صورة أكثر ثراءً عن هوية المشــاهدين، ومــاذا يريدون، وكيف يمكنك الحفاظ على انتباههم في سوق الترفيه المزدحم بشكل متزايد".

تقسيم المشاهدات:

يتمثل التحدي في أن هناك الكثير من المعلومات المتاحة فقط - مئات المليارات من صفوف البيانــات من مزودي خدمات بيانات الصناعة مثــل Nielsen وcom-Score،

من قنوات مثل AMC's TV Everywhere، التي تعيش على شبكة الإنترنت وخدمات الفيديو حسـب الطلب، من شركاء البيـع بالتجزئة مثل iTunes وAmazon، وخدمات الفيديو عبر الإنترنت من جهات خارجية مثل Netflix and Hulu.

يقـول Vitaly Tsivin: "لا يمكننـا الاعتـماد على ملخصات عالية المسـتوى؛ فنحن نحتـاج إلى أن نكون قادرين على تحليـل كل من البيانات المنظمة وغير المهيكلة، دقيقة بدقيقة ومشاهد بمشاهد"، ويضيف قائلًا: "نحتاج إلى معرفة من يشاهد المحتوى ولماذا، ونحتاج إلى معرفة ذلك بسرعة حتى نتمكن، على سبيل المثال، من اتخاذ قرار بشأن ما إذا كان سيتم عرض إعلان أو عرض ترويجي في موضع معين خلال حلقة الرجال المجانين".

وقـد قررت شركة AMC أنها بحاجة إلى تطويـر القدرة التحليلية في مجال الصناعة داخـل الشركة -وتركز على توفير هـذه الإمكانية في أسرع وقت ممكن. وبدلًا من إجراء عمليـة اختيار طويلـة ومكلفة للموردين والمنتجين، قررت شركة AMC الاسـتفادة من علاقتها الحالية مع شركة IBM كشريكها التقني الإستراتيجي الموثوق به. وبدلًا من ذلك، فقد تم استثمار الوقت والمال الذي يتم إنفاقه تقليديًّا على الشراء في تحقيق الحل -مما يسرع من تقدم AMC على خريطة الطريق الخاصة بتحليلاتها لمدة ٦ أشهر على الأقل.

تمكين قسم البحوث:

لقـد قـضى فريق البحـث في AMC في الماضي جـزءًا كبيرًا من الوقـت في معالجة البيانـات. واليوم تمكنـت AMC من تركيز معظـم طاقتها على اكتسـاب رؤى قابلة للتطبيق، وذلك بفضل أدوات التحليل الجديدة.

يقول Vitaly Tsivin: "من خلال الاسـتثمار في تقنيـة تحليلات البيانات الضخمة مـن IBM، تمكنا من زيـادة وتيرة أبحاثنا وتفاصيلها، فالتحليلات التي كانت تسـتغرق أيامًا وأسابيع يمكن إنجازها الآن في دقائق، أو حتى ثوانٍ. إن توفير التحليلات الداخلية سـيوفر الكثير مـن التكاليف. فبدلًا مـن دفع مئات الآلاف من الـدولارات إلى الباعة الخارجيين عندما نحتاج إلى بعض التحليل، فإنه يمكننا القيام بذلك بأنفسنا -بدقة أكثر، وسرعـة كبيرة، وتكلفة أقل بكثير. ونحن نتوقع رؤية عائد سريع على الاسـتثمار. ومع توفر المزيد من مصادر الرؤية المحتملة والتحليلات التي أصبحت أكثر إستراتيجية لقطاع الأعمال، فإن النهج الداخلي هو في الواقع الطريقة الوحيدة المقبولة لأي شـبكة ترغب حقًا في اكتساب ميزة تنافسية من بياناتها".



توجه القرارات بالبيانات:

تُظهـر العديد من النتائج التي توصل إليها هذا التحليـل الجديد تحولًا حقيقيًا في طريقة عمل AMC. فعلى سـبيل المثال، نجد أن قسم ذكاء الأعمال في الشركة استطاع إنشـاء نماذج إحصائية متطورة تسـاعد الشركة على تحسـين إسـتراتيجياتها التسويقية واتخاذ قرارات أكثر ذكاءً بشأن مدى قوة الترويج لكل عرض.

مجهزة	تقـوم شركة AMC بجمـع بيانات التقييمات مع معلومات المشـاهد من مجموعـة كبيرة مـن القنوات الرقمية: خدمـات الفيديو الخاصة بهـا عنـد الطلب وخدمـات البث المباشر وتجـار التجزئـة وخدمات التلفاز عبر الإنترنت.	
مترابطة	يعمل محرك البيانـات والتحليلات الضخـم والشـامل عـلى تحفيـز البيانات وجعلهـا متاحة لمجموعة مـن الأدوات التحليليـة الوصفيـة والتنبؤية للنمذجة السريعة والإبلاغ والتحليل السريع.	

ذكية	يستطيع AMC التنبؤ بالبرامج الناجحة، وكيفية ترتيبها، والعروض التي يجب أن تنشئها، ومن يجب عليها تسويقها- مما يساعد على كسب جماهير جديدة في سوق المنافسة المتزايدة.	

وبنظرة أكثر عمقًا على نسبة المشاهدة، فإن حملات AMC التسويقية المباشرة تحقق هي الأخرى نجاحًا أكبر. وفي أحد الأمثلة الحديثة، ساعدت التجزئة الذكية ونمذجتها الشبيهة الشركة في استهداف المشاهدين الجدد والحاليين بشكل فعال بحيث كانت معاملات الفيديو حسب الطلب في AMC أعلى مما كان متوقعًا بخلاف ذلك.

والجدير بالذكر أن هذه القدرة المكتشفة حديثًا للوصول إلى مشاهدين جدد والتي تستند إلى احتياجاتهم وتفضيلاتهم الفردية لا تُعد فقط قيمة بالنسبة إلى AMC - بل إنها تتمتع أيضًا بقيمة هائلة محتملة لشركاء الإعلان في الشركة. حيث تعمل AMC حاليًا على توفير الوصول إلى مجموعات البيانات وأدوات التحليل الغنية الخاصة بها كخدمة للمعلنين، مما يساعدهم على تحسين حملاتهم من أجل جذب جماهير أكبر من ذي قبل عبر القنوات الخطية والرقمية.

ويخلص Vitaly Tsivin إلى أنه: «يمكننا الآن الاستفادة من قيمة البيانات الكبيرة، يمكننا بناء اقتراح أكثر جاذبية لكل من المستهلكين والمعلنين، وإنشاء محتوى أفضل وتسويقه بشكل أكثر فاعلية ومساعدته على الوصول إلى جمهور أوسع من خلال أخذ الاستفادة الكاملة من قدراتنا متعددة القنوات».

أسئلة للمناقشة:

١- ما هي التحديات الشائعة التي تواجهها شركات البث في الوقت الحاضر؟ وكيف يمكن أن تساعد التحليلات على التخفيف من هذه التحديات؟

٢- كيف استفادت AMC من التحليلات لتعزيز أداء أعمالها؟

٣- ما هي أنواع تحليلات النصوص والحلول المصغرة النصية التي طورتها شبكات AMC؟ وهل يمكنك التفكير في استخدامات أخرى محتملة لتطبيقات استخراج النص في صناعة البث؟

Sources: IBM Customer Case Study. Using analytics to capture new viewers, predict ratings and add value for advertisers in amultichannel world. http://www-03.ibm.com/software/busi-nesscasestudies/us/en/corp?synkey=A023603A76220M60 (accessed July 2016); www.ibm.com; www.amcnetworks.com.

ويتـم تطبيق معالجة اللغـات الطبيعية NLP بنجاح على مجموعـة متنوعة من المجالات لمجموعة واسعة من المهام عبر برامج الحاسب لمعالجة اللغة البشرية الطبيعية تلقائيًا والتي كان يمكن القيام بها في السابق من قِبل البشر فقط. وفيما يلي بعضٌ من أكثر هذه المهام شيوعًا:

- **إجابة الأسـئلة:** وهي مهمة الرد التلقائي على السؤال المطروح باللغة الطبيعية؛ أي إنتاج إجابة لغويـة عند طرح سـؤال لغوي بشري. وللعثور على إجابة لسـؤال، فإن برنامج الحاسـب قد يسـتخدم قاعدة بيانات تم تنظيمها أو مجموعة من وثائق اللغة الطبيعية (نص أسـاسي مثل World Wide Web).
- **التلخيص التلقائي:** وهو إنشاء نسخة من وثيقة نصية بواسطة برنامج حاسب يحتوي على أهم النقاط في الوثيقة الأصلية.
- **توليد لغة طبيعية:** حيث تقوم الأنظمة بتحويل المعلومات من قواعد بيانات الحاسب إلى لغة بشرية يمكن قراءتها.
- **فهـم اللغة الطبيعية:** حيث تقوم الأنظمـة بتحويل عينات من اللغة البشرية إلى تمثيلات أكثر رسمية يسهل على برامج الحاسب التلاعب بها.
- **الترجمة الآلية:** وهي الترجمة التلقائية للغة إنسانية واحدة إلى لغة أخرى.
- **قراءة اللغة الأجنبية:** برنامج حاسـوبي يسـاعد المتحدث باللغات غـير الأصلية على قراءة لغة أجنبية مع النطق الصحيح واللهجات على أجزاء مختلفة من الكلمات.
- **الكتابة باللغة الأجنبية:** برنامج حاسب يساعد المستخدم غير الأصلي في الكتابة بلغة أجنبية.
- **التعرف على الكلام:** يحول الكلمات المنطوقة إلى مدخلات قابلة للقراءة آليًا. وبالنظر إلى مقطع صوت لشخص يتحدث، ينتج النظام إملاءً نصيًا.
- **النـص إلى الكلام:** يطلـق أيضًا عليه تركيب الكلام، يقوم برنامج الحاسـب تلقائيًا بتحويل نص اللغة العادية إلى خطاب إنساني.
- **تدقيق النص:** يقوم برنامج الحاسب بقراءة نسخة من النص لإثبات وتصحيح أي أخطاء.
- **التعـرف الضوئي على الحروف:** الترجمة الآلية لصور مكتوبة بخط اليد أو مطبوعة أو مطبعية (عادةً ما يتم التقاطها بواسطة ماسح ضوئي) وتحويلها إلى وثائق نصية قابلة للتحرير آليًا.

ويعتمد نجاح وشــهرة التنقيب في النص على التقــدم في معالجة اللغات الطبيعية NLP في كل جيــل كما هو الحال في فهــم اللغات البشرية. وتتيــح معالجة اللغات الطبيعية NLP اســتخراج الخصائــص من النص غير المهيكل بحيث يمكن اســتخدام مجموعة متنوعة من تقنيات التنقيب في البيانات لاستخراج المعرفة (أنماط وعلاقات جديدة ومفيدة) من ذلك النص.

أسئلة مراجعة على القسم ٥-٣:

١- ما هي معالجة اللغات الطبيعية NLP؟

٢- كيف ترتبط معالجة اللغات الطبيعية بالتنقيب في النصوص؟

٣- اذكر بعض الفوائد والتحديات لمعالجة اللغات الطبيعية NLP؟

٤- ما هي المهام الأكثر انتشارًا التي تتناولها معالجة اللغات الطبيعية NLP؟

٥-٤ تطبيقات التنقيب في النص:

نظرًا لزيادة كمية البيانات غير المنظمة التي يتم جمعها من قبل المؤسسات، فإن اقتراح أدوات للتنقيب في النص فائقة القيمة وذات انتشــار واســع تزداد أيضًا. وتدرك العديد من المنظمات في الوقت الراهن أهمية اســتخراج المعرفة من مســتودعات البيانات المستندة إلى الوثائق من خلال اســتخدام أدوات التنقيــب في النــص. وفيما يلي نذكــر مجموعة فرعية صغــيرة فقط من فئات التطبيقات المثالية للتنقيب في النص.

تطبيقات التسويق:

يمكن اســتخدام التنقيب في النص لزيادة البيع المتقاطع والبيع من خلال تحليل البيانات غير المهيكلة الناتجة عن مراكز الاتصال. كما يمكن تحليل النصوص التي تم إنشاؤها بواسطة ملاحظات مركز الاتصال، بالإضافة إلى نسخ المحادثات الصوتية مع العملاء من خلال خوارزميات التنقيب في النص لاستخراج معلومات جديدة وعملية حول تصورات العملاء تجاه منتجات وخدمات الشركة. وبالإضافة إلى ذلك، تعد المدونات، وملاحظات المســتخدمين للمنتجات في مواقع الويب المستقلة، ومنشورات لوحات المناقشة كنزًا ذهبيًا يمكن من خلاله التعرف على مشاعر العملاء؛ حيث يمكن استخدام هذه المجموعة الغنية من المعلومات، بمجرد تحليلها بشكل صحيح، لزيادة رضا العميل والقيمة الإجمالية لمدة تعامله مع الشركة. (Coussement & Van den Poel, 2008).

لقــد أصبحــت قيمة التنقيب في النــص لا تُقدر بثمــن لإدارة علاقات العمــلاء؛ حيث يمكن للشركات اســتخدام التنقيب في النص لتحليل مجموعات غنية من البيانات النصية غير الهيكلية، مقترنة بالبيانات المنظمة ذات الصلة المســتخرجة من قواعد البيانات التنظيمية، للتنبؤ بتصورات العملاء وســلوك الشراء اللاحق. وقد نجــح كل مــن Coussement وVan den Poel (2009) في تطبيــق التنقيــب في النص بنجاح من أجل تحســين قدرة نموذجٍ ما على التنبؤ بشــكل أفضل باضطراب العملاء (أي تناقص العملاء) بحيث يتم تحديد العملاء الأكثر عرضة لمغادرة الشركة من أجل القيام بالتكتيكات اللازمة للحفاظ عليهم.

وقد اســتخدم Ghani وآخــرون (٢٠٠٦) التنقيب في النص لتطوير نظام قادر على اســتنتاج سمات ضمنية وصريحة من المنتجات من أجل تعزيز قدرة تجار التجزئة على تحليل قواعد بيانات المنتجات. إن معالجة المنتجات كمجموعات من أزواج الســمة والقيمة وليس ككيانات بســيطة يمكن أن تعزز فعالية العديد من تطبيقات الأعمال، بما في ذلك التنبؤ بالطلب، وتحســين التركيب، بالإضافــة إلى توصيات المنتجات، ومقارنة التركيب بــين تجار التجزئة والمصنعين، فضلًا عن اختيار موردي المنتجات. ويســمح النظام المقترح للأعمال التجارية بتمثيل منتجاتها من حيث الســمات وقيم السمات بدون جهد يدوي كبير. كما يتعلم النظام هذه السمات عن طريق تطبيق تقنيات التعلم تحت إشراف شــبه موجــه semisupervised على وصف المنتجــات الموجودة في مواقع الويب الخاصة بتجار التجزئة.

تطبيقات الأمان:

ربما يكون أحد أكبر وأبرز تطبيقات التنقيب في النص في مجال الأمان نظام المراقبة ECHELON المصنــف بدرجة عالية. وكما هو معروف، فإنه مــن المفترض أن يكون نظام ECHELON قادرًا عــلى تحديد محتوى المكالمات الهاتفية والفاكســات والبريد الإلكتروني وأنواع أخرى من البيانات، بالإضافة إلى اعتراض المعلومات المرسلة عبر الأقمار الصناعية، وشبكات الهاتف العمومية التبديلية، وروابط الميكروويف.

وفي عام ٢٠٠٧م، قامت EUROPOL بتطوير نظام متكامل يمكنه الوصول إلى كميات هائلة من مصادر البيانات المهيكلة وغير المهيكلة، وتخزينها وتحليلها لتتبع الجريمة المنظمة عالميًا. ويطلق على نظام التحليل الشامل لدعم أجهزة الذكاء اسم نظام (OASIS)، ويهدف هذا النظام إلى دمج أحدث تقنيات التنقيب في البيانات والنص المتوفرة في السوق في الوقت الحالي. وقد مكن نظام EUROPOL من إحراز تقدم كبير في دعم أهداف تطبيق القانون على المستوى الدولي (EUROPOL، 2007).

ويقوم كل من مكتب التحقيقات الفيدرالي الأمريكي (FBI) ووكالة المخابرات المركزية (CIA)، تحت إشراف وزارة الأمن الوطني، بتطوير نظام للبيانات العملاقة والنصوص. ومن المتوقع أن يُنشئ النظام مستودع بيانات ضخمٍ إلى جانب مجموعة متنوعة من وحدات البيانات والتنقيب في النص لتلبية احتياجات اكتشاف المعرفة لوكالات تطبيق القانون الفيدرالية والولائية والمحلية. وقد كان لكل من مكتب التحقيقات الفيدرالي ووكالة المخابرات المركزية قبل هذا المشروع قاعدة بيانات منفصلة خاصة بها، مع القليل من الترابط أو حتى بدونه.

وهناك تطبيق آخر من تطبيقات التنقيب في النصوص، هذا التطبيق في مجال كشف الخداع، وهو تطبيق التنقيب في النص على مجموعة كبيرة من التصريحات الإجرامية في العالم الحقيقي (شخصية الفائدة)، وقد وضع كل من Fuller، وBiros، وDelen (2008) نماذج التنبؤ لتمييز البيانات الكاذبة من البيانات الصادقة. وباستخدام مجموعة غنية من الإشارات المستخلصة من البيانات النصية، يتنبأ النموذج بعينات الدقة بنسبة ٧٠٪، والتي يعتقد أنها نجاح كبير مع الأخذ في الاعتبار أن الإشارات يتم استخلاصها فقط من البيانات النصية (لا توجد إشارات شفوية أو بصرية). علاوةً على ذلك، فبالمقارنة مع تقنيات كشف الخداع الأخرى، مثل جهاز كشف الكذب، فإن هذه الطريقة غير متداخلة وقابلة للتطبيق على نطاق واسع ليس فقط للبيانات النصية، ولكن أيضًا (محتمل) إلى نسخ التسجيلات الصوتية. ويرد وصف أكثر تفصيلًا للكشف عن الخداع القائم على النص في الحالة العملية ٥-٣.

حالة عملية ٥-٣

التنقيب للكشف عن الأكاذيب

تواصل الاتصالات المبرمجة تقدمها مدفوعة بالتطورات في تقنيات المعلومات على شبكة الإنترنت وزيادة العولمة، لتتحول إلى الحياة اليومية، حاملة معها أساليب جديدة للخداع. فنجد أن هناك تزايدًا سريعًا في حجم الدردشة النصية، والرسائل الفورية، والرسائل النصية، بالإضافة إلى النص الذي تولده مجتمعات الإنترنت. بل حتى البريد الإلكتروني هو أيضًا مستمر في تزايد استخدامه. ومع النمو الهائل في التواصل القائم على النصوص، ازدادت إمكانات الناس لخداع الآخرين من خلال التواصل المحوسب، مما قد يؤدي إلى نتائج كارثية.

ولسوء الحظ، فإن البشر بشكلٍ عام، يميلون إلى التكاسل وعدم بذل الجهد في مهام كشف الخداع، وتتفاقم هذه الظاهرة في الاتصالات القائمة على النص. وقد شمل جزء كبير من البحث عن كشف الخداع (المعروف أيضًا بتقييم المصداقية) في الاجتماعات

والمقابلات الشــخصية المباشرة. ومع ذلك، ومع نمو الاتصالات القائمة على النصوص، فإن تقنيات الكشف عن الخداع المستندة إلى النصوص تُعد ضرورية.

وهناك تقنيات للكشــف عن الخــداع بنجاح، وقد حظيت هــذه التقنيات بقابلية تطبيقها على نطاق واســع؛ حيث يمكن لتطبيق القانون اســتخدام أدوات وتقنيات دعم القــرار للتحقيــق في الجرائم، وإجــراء الفحص الأمنــي في المطــارات، ومراقبة اتصالات الإرهابيين المشــتبه بهم. وقد يســتخدم المتخصصون في الموارد البشرية أدوات كشــف الخداع لفحص المتقدمين. فهذه الأدوات والتقنيات لديها أيضًا القدرة على فحص رسائل البريد الإلكتروني للكشــف عن الاحتيال أو غيرها من الأفعال الخاطئة التي ارتكبها ضباط الشركات. وعلى الرغم من أن بعض الناس يعتقدون أنهم يســتطيعون التعرف بســهولة على أولئك الذين ليســوا صادقين، إلا أن ملخصًا لأبحاث الخداع أظهر أن متوســط نسبة الأشخاص الذين يدققون في اتخاذ قرارات صائبة هي ٥٤٪ (Bond & DePaulo, 2006)، وقد يكون هذا الرقم في الواقع أسوأ عندما يحاول البشر اكتشاف الخداع في النص.

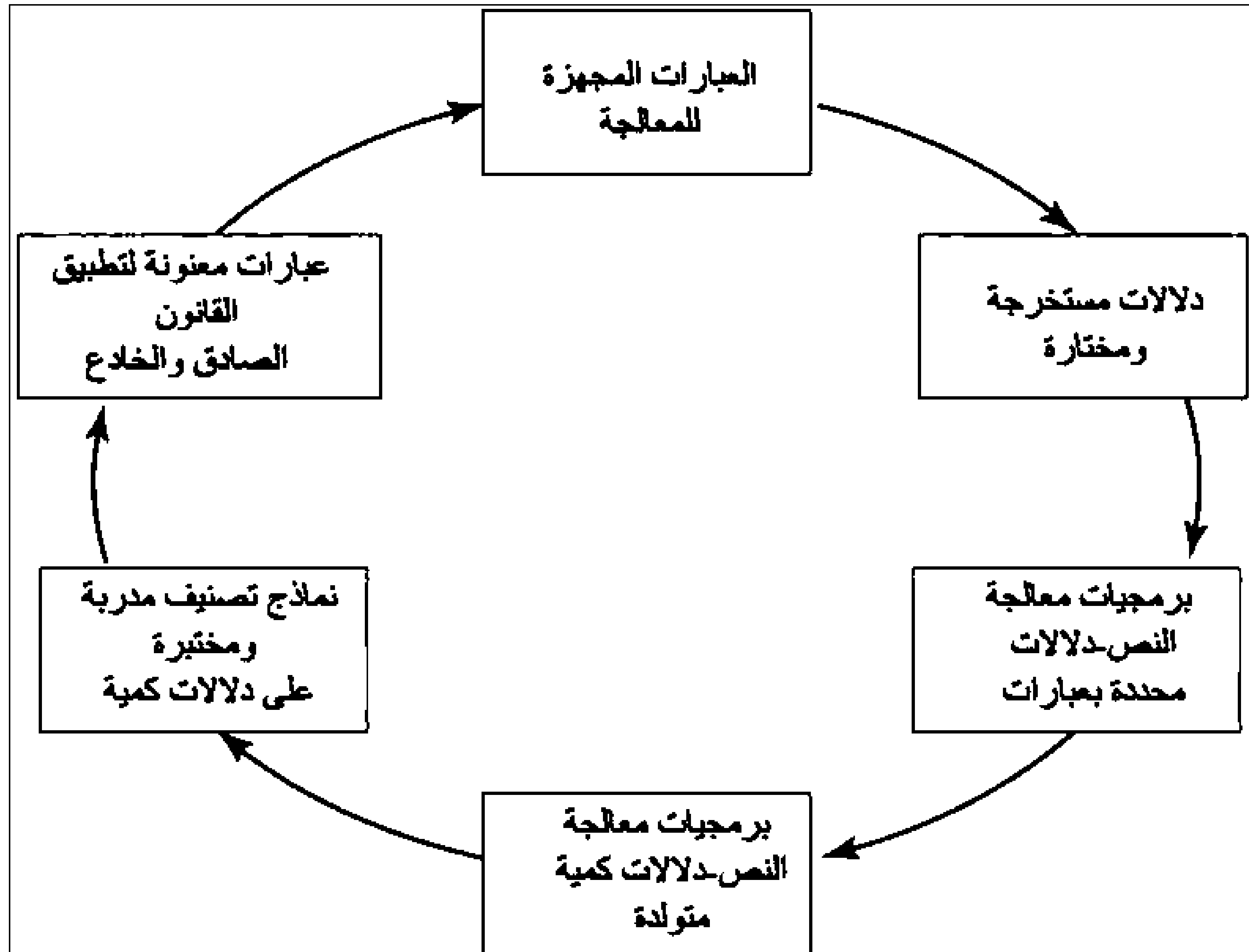
وباستخدام مجموعة من تقنيات التنقيب في النص والتنقيب في البيانات، قام كل من Fuller وآخرون (٢٠٠٨م) بتحليل البيانات الشــخصية التي استغلها أشخاص متورطون في جرائم في قواعد عســكرية؛ حيث يُطلب من المشــتبه بهم والشهود في هذه البيانات كتابة ما يتذكرونه مــن الحدث بكلماتهم الخاصة. وقد قام الموظفون المكلفون بتنفيذ القانون العســكري بتفتيش البيانات الأرشــيفية للإدلاء بتصريحات يمكن أن يحددوها بشــكل قاطع بأنها صادقــة أو خادعة. وقد اتُخذت هذه القرارات على أســاس الأدلة الموثقــة وحالة القضية. وبعــد أن يتم تصنيفهم كصادقين أو مخادعين، يقوم الموظفون المكلفــون بتنفيذ القانون بإزالة معلومات تحديد الهويــة ويقدمون البيانات إلى فريق البحــث. وفي المجمــوع، تم تلقي ٣٧١ بيانًا قابلًا للاســتخدام للتحليــل. وقد اعتمدت طريقة الكشف عن الخداع المبنية على النص التي استخدمها Fuller وآخرون (٢٠٠٨م)، على عملية تُعرف باســم التنقيب في خاصية الرسائل، والتي تعتمد على عناصر البيانات وتقنيات التنقيب في النص. ويعرض الشكل (٥-٣) وصفًا مبسطًا للعملية.

أولًا، أعد الباحثون البيانات للمراجعة، حيث يجب أن يتم نســخ البيانات الأصلية المكتوبــة بخــط اليــد في ملف معالجــة النصوص. ثانيًــا، تم تحديد الســمات (مثل: الإشارات)، حيث قام الباحثون بتحديد ٣١ سمة تمثل فئات أو أنواعًا من اللغة مستقلة نســبيًا عن محتوى النص ويمكن تحليلها بســهولة عن طريق وسائل آلية. فعلى سبيل

المثال، يمكن تعريف ضمير الشـخص الأول، مثل: أنـا أو لي دون تحليل للنص المحيط. ويسرد الجدول (٥-١) الفئات وقائمة أمثلة بالخصائص المستخدمة في هذه الدراسة.

وقد تم اسـتخراج الميـزات من البيانات النصية وإدخالهـا في ملف ثابت لمزيد من المعالجة. وباستخدام العديد من طرق اختيار الخصائص بالإضافة إلى التحقق المتقاطع من ١٠ أضعاف، قارن الباحثون دقة التنبؤ بثلاث طرق شائعة للتنقيب في البيانات. وقد أشـارت النتائج إلى أن نماذج الشبكات العصبية كانت صاحبة الأداء الأفضل بنسبة دقة تنبؤية بلغت ٧٣,٤٦٪ على عينات بيانات الاختبار، وجاءت أشجار القرار في المركز الثاني بنسبة ٧١,٦٪، في حين حل في المركز الأخير الانحدار اللوجستي بنسبة ٦٥,٢٨ ٪.

وتشير النتائج إلى أن الكشف التلقائي عن الخداع المبني على النص لديه القدرة على مسـاعدة المكلفين بهذه المهمة، ويمكن تطبيق هذه الطريقـة بنجاح على بيانات العالم الحقيقي. وقد تجاوزت دقة هذه التقنيات دقة معظم تقنيات كشـف الخداع الأخرى، وذلك على الرغم من أنها كانت تقتصر على الإشارات النصية.



شكل ٥-٣: عملية كشف الخداع القائمة على النص

جدول ٥-١: فئات وأمثلة من الميزات اللغوية المستخدمة في كشف الخداع

م	الفئة	إشارات المثال
١	الكمية	عدد الأفعال، عدد عبارات الأسماء، إلخ
٢	التعقيد	متوسط عدد الجمل ومتوسط طول الجملة وما إلى ذلك
٣	عدم اليقين (الشك)	المعدلات، الأفعال الشرطية، إلخ
٤	عدم الفورية	الصوت المنفعل، والتصنيف، وما إلى ذلك
٥	التعبيرية	Emotiveness
٦	التنوع	التنوع المعقول، التكرار، إلخ
٧	الغير رسمية	نسبة الأخطاء المطبعية
٨	الخصوصية	المعلومات الزمانية المكانية، والمعلومات الإدراكية، وما إلى ذلك
٩	التأثير	تأثير إيجابي، وتأثير سلبي، إلخ

أسئلة للمناقشة:

١- لماذا يصعب كشف الخداع؟

٢- كيف يمكن استخدام التنقيب في النص / البيانات للكشف عن الخداع في النص؟

٣- ماهي التحديات الرئيسة في رأيك لمثل هذا النظام الآلي؟

Sources: Fuller, C. M., Biros, D., & Delen, D. (2008). Exploration of feature selection and advanced classification models for high-stakes deception detection. Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS), Big Island, HI: IEEE Press, 8099-; Bond C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. Personality and Social Psychology Reports, 10(3), 214234-.

التطبيقات الطبية الحيوية:

يحمـل التنقيب في النص إمكانات كبيرة للمجال الطبي بشـكل عام والطب الحيوي على وجه الخصوص، وذلك لعدة أسـباب. أولًا، التوسـع في الأدبيات (المواد المطبوعة) والمنشورات المنشورة (خاصـة مـع ظهور المجلات المفتوحة المصـدر) في الميدان بمعدل أُسي. ثانيًـا، بالمقارنة مع معظم

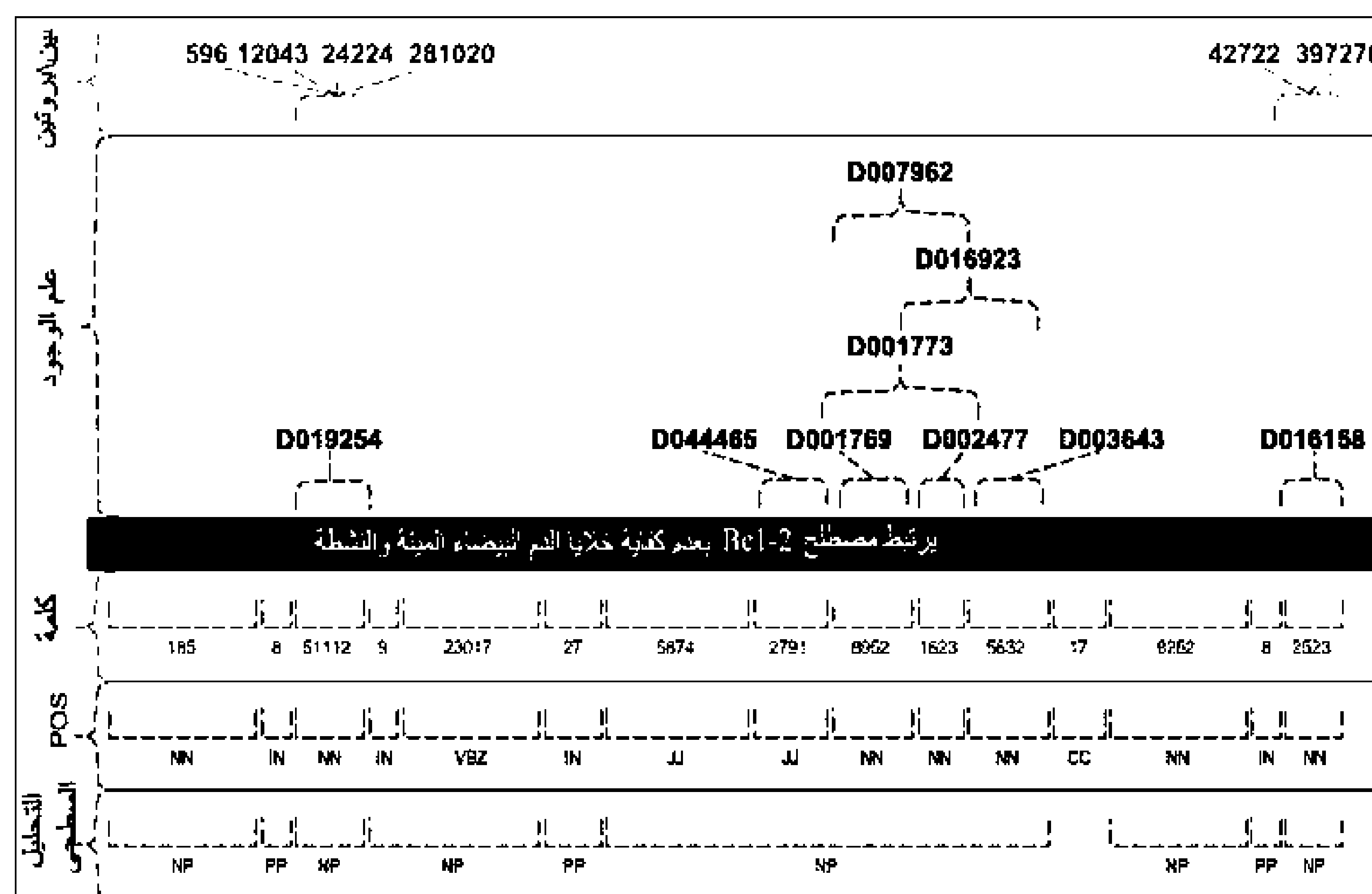
المجـالات الأخرى، فإن الأدبيات (المـواد المطبوعة) الطبية هي أكثر توحيـدًا وتنظيمًا، مما يجعلها مصدر معلومات "أكثر ملاءمة". وأخيرًا، تكون المصطلحات المستخدمة في هذا الكتاب ثابتة نسبيًا، مع وجود (علم الوجود) ontology الموحد إلى حدٍ ما. وفيما يلي بعض الدراسات النموذجية حيث تم استخدام تقنيات التنقيب في النص بنجاح في استخراج أنماط جديدة من الأدب الطبي الحيوي.

إن التقنيـات التجريبية مثل تحليل الحمض النووي المجهري (DNA)، والتحليل المتسلسـل للانطباع الجيني (SAGE)، ومجموعة spectrometry proteomics، وغيرها، تولد كميات كبيرة مـن البيانات المتعلقـة بالجينات والبروتينات. وكـما هو الحال في أي نهـج تجريبي آخر، فمن الضروري تحليل هذا الكم الهائل من البيانات في سياق المعلومات المعروفة سابقًا حول الكيانات البيولوجيـة قيد الدراسـة. وتعتبر الأدبيـات مصدرًا قيمًا للغاية للمعلومـات للتحقق من صحة التجارب وتفسيرها. ولذلك؛ فإن تطوير أدوات التنقيب الآلي عن النصوص للمساعدة في مثل هذا التفسير هو أحد التحديات الرئيسية في أبحاث المعلوماتية الحيوية الحالية.

إن معرفـة موقـع البروتـين داخل الخلية يمكنه أن يسـاعد على توضيـح دور تلك الخلية في العمليـات الحيوية وتحديـد إمكاناتها كهدف دوائي. وقد تم وصـف العديد من أنظمة تحديد الموقـع في الأدبيات؛ حيـث يركز البعض على كائنات معينة، في حين يحـاول البعض الآخر تحليل مجموعة واسعة من الكائنات الحية. وقد اقترح Shatkay وآخرون (٢٠٠٧م) نظامًا شاملًا يستخدم عدة أنواع من السـمات المتسلسـلة والنصية للتنبؤ بموقع البروتينـات. ويكمن التجديد الرئيس لنظامهم في الطريقة التي يتم بها اختيار مصادر النص وميزاته ودمجها مع ميزات متسلسلة. وقد قاموا باختبار النظام على مجموعات البيانات المسـتخدمة سابقًا وعلى مجموعات بيانات جديدة وضعـت خصيصًا لاختبار قوتها التنبؤية، حيث أظهرت النتائج أن نظامهم يتغلب باسـتمرار على النتائج التي تم الإبلاغ عنها مسبقًا.

أمـا Chun وآخرون (٢٠٠٦م) فقد قاموا بوصف نظام يسـتخرج علاقـات الجين المرضي من الأدبيـات التـي يتم الوصول إليها عـبر MEDLINE، حيث قاموا ببناء قاموس لأسـماء الأمراض والجينات من سـت قواعد بيانات عامة واسـتخرجوا العلاقة بين المرشـحين عـن طريق مطابقة القامـوس. ونظرًا لأن مطابقة القاموس تنتج عددًا كبيرًا من النتائج الإيجابية لكنها غير صحيحة، فقـد طوّروا طريقـة للتعرُّف على الكيان القائم على تعلم الآلة المسـمى بكيان التعرف (NER) لتصفيـة التقديرات الخاطئة لأسـماء الأمـراض / الجينات. ووجد الباحثون أن نجاح اسـتخلاص العلاقة يعتمد بشـكل كبير على أداء تصفية NER وأن التصفية حسّـنت دقة اسـتخلاص العلاقة بنسبة ٢٦,٧٪، على تكلفة انخفاض صغير في إعادة الاتصال.

ويبين الشـــكل (٥-٤) صورة مبسطة لعملية تحليل نص متعدد المستويات لاكتشاف العلاقات بـــين الجينـــات والبروتينات (أو تفاعـــلات البروتينـــات والبروتينات) في الأدبيـــات الطبية الحيوية (Nakov وآخرون، ٢٠٠٥). وكما يتبين في هذا المثال المبســـط الذي يســـتخدم جملة بســـيطة من النص الطبي الحيوي، حيث يتم أولًا (في المســـتويات الثلاثة السفلية) ترميز النص باستخدام جزء مـــن الكلام مع قليل مـــن التعديل، ثم تتم مطابقة المصطلحات (الكلمات) الرمزية (وتفســـيرها) مقابل التمثيل الهرمي لنطاق علم الوجود لاشـــتقاق العلاقة بين الجين والبروتين. إن تطبيق هذه الطريقة (و / أو بعض اختلافاتها) على الأدبيات الطبية الحيوية يوفر إمكانات كبيرة لفك شـــفرة التعقيدات في مشروع Human Genome.



شكل ٥-٤: تحليل متعدد المستويات للنص لتحديد تفاعل الجين/ البروتين

التطبيقات الأكاديمية:

تحظى مســـألة التنقيب في النص بأهمية كبيرة بين الناشرين الذين لديهم قواعد بيانات كبيرة من المعلومات التي تتطلب فهرســـة لاســـترجاعها بشـــكل أفضل. وهذا صحيح بشـــكل خاص في التخصصات العلمية، والتي غالبًا ما تكون معلومات محددة للغاية يتضمنها النص المكتوب. وقد تـــم إطلاق مبـــادرات، مثل: اقتراح الطبيعة لواجهة التنقيب في النص المفتوح، والتعريف الشـــائع لنوع وثيقة نشر المعاهد الوطنية للصحة، والتي من شـــأنها أن توفر مؤشرات دلالية آلية للإجابة عن استفسارات محددة موجودة في النص دون إزالة حواجز الناشر للوصول إلى الجمهور.

كما أطلقت المؤسسـات الأكاديمية مبادرات التنقيب في النص. فنجد على سبيل المثال أن المركز الوطنـي للتنقيـب في النـص - وهو جهد تعاوني بين جامعتَي مانشسـتر وليفربـول - يوفر أدوات مخصصة، ومرافق بحث، ومشورة حول التنقيب في النص للمجتمع الأكاديمي، مع التركيز المبدئي على التنقيب على النص في العلوم البيولوجية والطبية الحيوية، وقد توسعت الأبحاث منذ ذلك الحين في العلـوم الاجتماعية. وفي الولايات المتحدة، تقوم كليـة الإعلام في جامعة كاليفورنيا، بيركلي، بتطوير برنامج يدعى bioText لمساعدة الباحثين في العلوم البيولوجية في التنقيب في النصوص وتحليلها.

وكما هو موضح في هذا القسـم، فإن التنقيب في النص يحتوي على مجموعة متنوعة من التطبيقات في عدد من التخصصات المختلفة. انظر الحالة العملية (٥-٤) للاطلاع على مثال لكيفية استخدام الشركة الرائـدة في تصنيع منتجات الحوسـبة في معالجة النصـوص؛ من أجل فهم احتياجـات ورغبات العملاء الحاليين والمحتملين المتعلقة بجودة المنتج وتصميم المنتج بشكل أفضل.

حالة عملية ٥-٤

إدراج العميـل ضمن معادلة الجودة: تسـتخدم لينوفو Lenovo التحليلات من أجل إعادة التفكير في إعادة تصميمها

كانـت لينوفو Lenovo تقترب من التصميم النهائي لتحديث تخطيط لوحة المفاتيح لأحد أشهر أجهزة الحاسـب الخاصة بها عندما رصدت مجتمعًا صغيرًا من اللاعبين على الإنترنـت، غير أنه يعتبر مجتمعًا مهـمًا وكبيرًا، يدعم تصميم لوحة المفاتيح الحالي. وقد أدى هـذا التغيـير في التصميم إلى تمرد جماعي لقطاع كبير مـن قاعدة عملاء لينوفو - المطورين المستقلين واللاعبين.

وقـد كانت وحـدة تحليلات الشركات تسـتخدم SAS كجزء من مـشروع الجودة الجوهرية. وبتتبع الويب، وغربلة البيانات النصية التي تشير إلى لينوفو، كشف التحليل عـن منتدى لم يكن معروفًا من قبل، أن أحد العملاء الحاليين كتب مراجعة متألقة من ست صفحات للتصميم الحالي، خاصة لوحة المفاتيح. وقد جذب هذا الاستعراض ٢٠٠٠ تعليـق! يقول Mohammed Chaara، مدير تحليل العملاء وVOC Analytics: "لم يكن الأمر شيئًا وقد وجدناه في المراجعات التقليدية لتصميم ما قبل الإنتاج".

وقـد كان هذا النوع من الاكتشـافات الذي عـزز التزامات الشركـة لنظام لينوفو للكشف المبكر (LED)، وعمل Chaara وفريق تحليلات الشركات الخاص به.

لم تنطلق شركة لينوفو، أكبر شركة لأجهزة الحاسب الشخصية والأجهزة اللوحية على مستوى العالم، لقياس مشاعر المدونين الغامضين أو اكتشاف المنتديات الجديدة. أرادت الشركة إحراز الجودة وتطوير وابتكار المنتجات من خلال دراسة البيانات -الخاصة بها ومن خارج الجدران الأربعة. يقول Chaara: "نحن نركز بشكل رئيسي على تحسين سلسلة التوريد، وفرص البيع / بيع منتج عن طريق إقناع الزبون، والتسعير والتعبئة للخدمات. وأي تحسينات نجريها في هذه المجالات تستند إلى الاستماع إلى العميل". وتوفر SAS الإطار "لإدارة كمية البيانات المجنونة" التي يتم إنشاؤها.

وقد انتشر نجاح المشروع كالنار في الهشيم داخل المنظمة. حيث خططت لينوفو في البداية لنحو ١٥ مستخدمًا، ولكن شفهيًا وصلت إلى ٣٠٠ مستخدم قاموا بتسجيل الدخول إلى لوحة المعلومات LED لعرض مرئي حول معنويات العملاء، والضمان، وتحليل مركز الاتصال.

وكانت النتائج مذهلة:

- تخفيض أكثر من ٥٠٪ من وقت اكتشاف المشكلة.
- تخفيض بنسبة من ١٠٪ إلى ١٥٪ في تكاليف الضمان من العيوب الخارجة عن الطبيعي.
- تخفيض بنسبة من ٣٠٪ إلى ٥٠٪ في مكالمات مركز الاتصال للمعلومات العامة.

النظر إلى الصورة الكبيرة:

تتميز الطرق التقليدية لقياس المشاعر وجودة الفهم بنقاط الضعف والتأخر الزمني:

- تعرض استطلاعات العملاء معلومات فقط من العملاء الراغبين في ذلك.
- غالبًا ما تأتي معلومات الضمان بعد أشهر من تسليم المنتج الجديد.
- قد يكون من الصعب فك شفرة الكثير من الأسباب لسخط العملاء ومشكلات المنتج.

وبالإضافة إلى ذلك، تتبع لينوفو منتجها وتعبئته ببرامج لا تنتجها، ويستخدم العملاء مجموعة متنوعة من الملحقات (محطات الإرساء وأجهزة الماوس) التي قد تكون أولا تكون منتجات لينوفو. ولزيادة المشكلة، تعمل الشركة في ١٦٥ دولة وتدعم أكثر من ٣٠ لغة، لذلك كانت الطرق اليدوية لتقييم التعليق متعارضة، واستغرقت الكثير من الوقت، ولم يكن بمقدورها التوسع في حجم التعليقات التي كانت تراها في وسائل التواصل الاجتماعي. ويجب أن يكون تحليل المشاعر قادرًا على إدراك الفروق الدقيقة

في اللغات المحلية. (على سـبيل المثال، طريقة وصف الأسـتراليين للأشـياء تختلف عن طريقة وصف الأمريكان).

وقد وفر الاكتشاف القائم على التحليل لمشكلة في محطات قاعدة الشحن ثاني فوز كبير لمبادرة LED الخاصة بـ Lenovo؛ حيث كان العملاء يتصلون بالدعم الفني ليقولوا إنهم يواجهون مشـاكل في الشاشة، أو أن الجهاز مغلق بشكل مفاجئ أو أن البطارية لا تشـحن. وقد ظهرت حسابات مماثلة على مواقع وسائل التواصل الاجتماعي. في بعض الأحيان، ولكن ليس دائمًا، ذكر العميل لإرساء السفن docking. لم يكن الأمر كذلك قبل أن تسـتخدم لينوفو SAS لتحليل مجموعة مكوّنات مراكز الاتصال ومشـاركات وسائل التواصل الاجتماعي التي ربطت كلمة إرسـاء docking بالمشـكلة، مما ساعد مهندسي الجودة على معرفة السبب الجذري وإصدار تحديث للبرنامج.

يقـول Chaara: "لقد تمكنا من جمع هذه التعليقات في غضون أسـابيع، وكانت تستغرق من ٦٠ إلى ٩٠ يومًا لأننا اضطررنا إلى انتظار عودة التقارير من الميدان". والآن يسـتغرق الأمر من ١٥ إلى ٣٠ يومًا فقط. وقد أدى هذا الانخفاض في وقت الاكتشـاف إلى انخفاض بنسـبة ١٠٪ إلى ١٥٪ في تكاليف الضمان لهذه المشكلات، والتي تبلغ حوالي ٢,١ مليار دولار سنويًّا، وهذه توفيرات كبيرة.

وعـلى الرغم من أن معلومات مركز الاتصال كانت حاسـمة، إلا أن مكون وسـائل التواصل الاجتماعي هو ما أغلق الصفقة. يقول Chaara: "مع تويتر وفيسبوك، وصف النـاس مـا كانوا يفعلونه في تلك اللحظة قائلين: لقد اسـتقرت الآلـة وحدث X. إنها خامدة وغير متحيزة وقوية للغاية".

وقـد تم العثور عـلى فكرة غير متوقعة عند تحليل ما يقوله العملاء أثناء تشـغيل أجهزة الحاسب الشخصية الخاصة بهم. كما أدركت لينوفو أن وثائقها لشرح منتجاتها، والضمانات، وما شـابه لم يكن واضحًا. قـال Chaara: "هناك تكلفة لكل مكالمة لمركز الاتصال. وبفضل الوثائق المحسنة، شهدنا انخفاضًا بنسبة من ٣٠٪ إلى ٥٠٪ في المكالمات الواردة للحصول على معلومات عامة".

الإشادة بالفوز وراء الخطوط الأمامية:

كان المشروع ناجحًا جدًا لدرجة أن Chaara عرضه على CEO. والهدف هو تكوين عرض لوحة معلومات لـ C-suite. يقول Chaara: "هذا هو مسـتوى التفكير من كبار التنفيذيين

لدينـــا، فهم يؤمنون بهذا. وبالإضافة إلى ذلك، ســتقوم مجموعــة Chaara بقياس نجاح هذا الجهد بشكل رسمي وتوسيعه لقياس قضايا مثل تجربة العملاء عند شراء منتج Lenovo".

وقد أدى تطبيق التحليلات في نهاية المطاف إلى فهم أكثر شــمولية لمفهوم الجودة. فالجودة ليســت مجرد جهاز حاســب يعمل بشكل صحيح. ولكنها تعني معرفة الناس كيفية اســتخدامه، والحصول على مســاعدة سريعة ودقيقة من الشركة، والحصول على مكونات غير تابعة لشركة لينوفو للعمل بشكل جيد مع الأجهزة، وفهم ما يعجب العملاء بشأن المنتج الحالي -بدلًا من إعادة تصميمه فقط لأن مصممي المنتجات يعتقدون أنها الشيء الصحيح الذي ينبغي القيام به. يقول Chaara: "لقد سمحت لنا SAS بالحصول على تعريف للجودة من وجهة نظر العميل".

أسئلة للمناقشة:

١- كيف اســتخدمت لينوفو تحليلات النصوص والتنقيب في النص لتحســين جودة وتصميم منتجاتها وفي النهاية تحسين رضا العملاء؟

٢- ما هي التحديات والحل المقترح والنتائج التي تم الحصول عليها؟

المصــادر: "إحضــار العميــل إلى معادلــة الجــودة (SAS® Visual Analytics, SAS® Contextual Analysis, SAS® Sentiment Analysis, SAS® Text Miner) - تحليل المشاعر والتحليلات المتقدمة تســاعد لينوفو على تحديد مشكلات الجودة ورغبات العملاء بشــكل أفضل". Copyright © 2016 SAS Institute Inc., Cary, NC, USA. أعيد طبعها بإذن. كل الحقوق محفوظة.

أسئلة مراجعة على القسم ٥-٤:

١- اذكر وناقش باختصار بعضًا من تطبيقات التنقيب في النص في مجال التسويق.

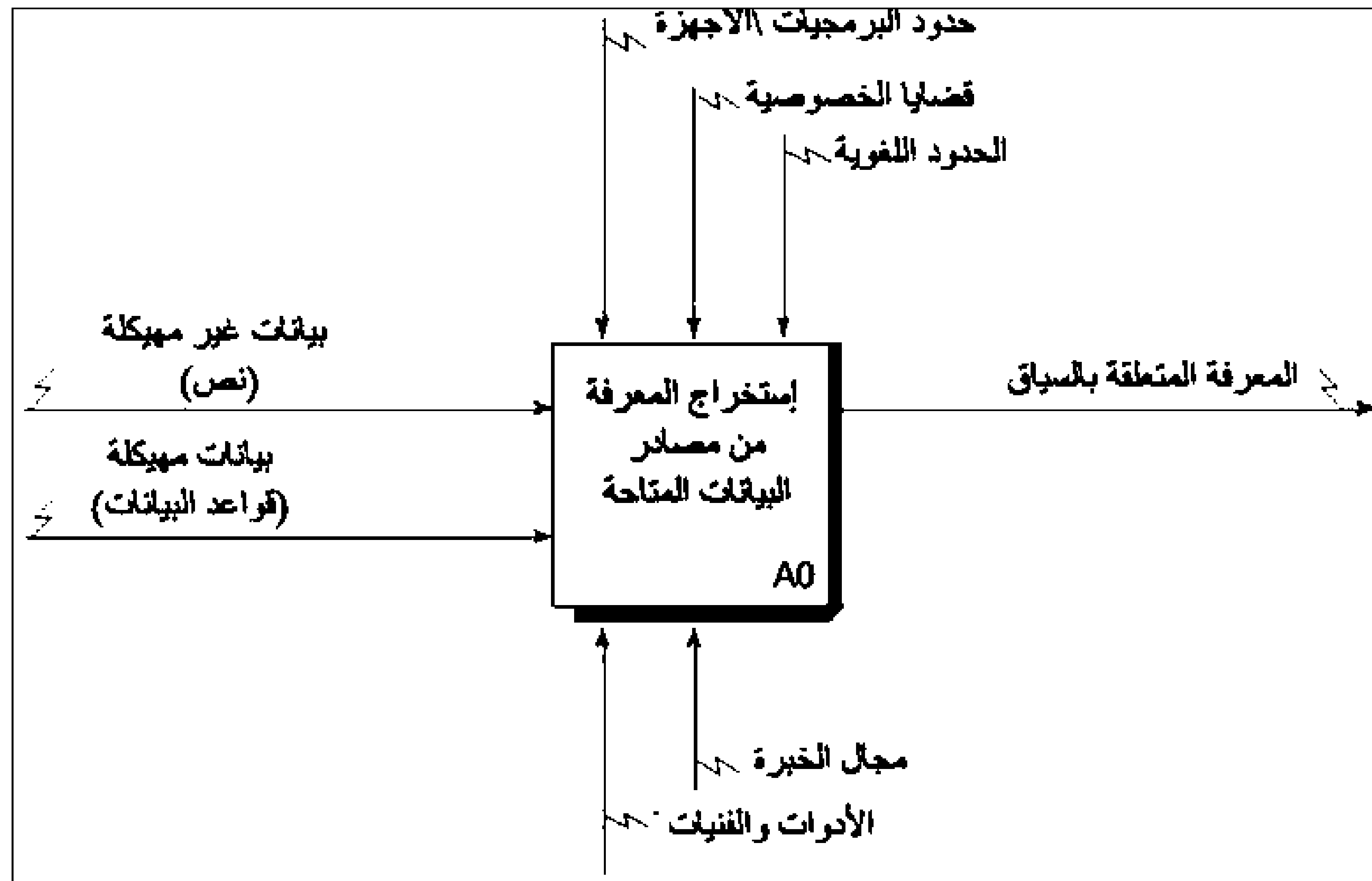
٢- كيف يمكن استخدام التنقيب في النص في الأمن ومكافحة الإرهاب؟

٣- ما هي بعض تطبيقات التنقيب الواعدة في النص في مجال الطب الحيوي؟

٥-٥ عملية التنقيب في النص:

لكي تنجح دراســات تحليل النص، فلابد أن تتبع منهجية ســليمة تستند إلى أفضل الممارسات. حيــث يلزم وجود نموذج عملية موحدة مماثل للعملية القياســية للتنقيب في البيانات الصناعية (CRISP-DM)، والذي يُعد معيار الصناعة لمشاريع التنقيب في البيانات (انظر الفصل ٤). وعلى

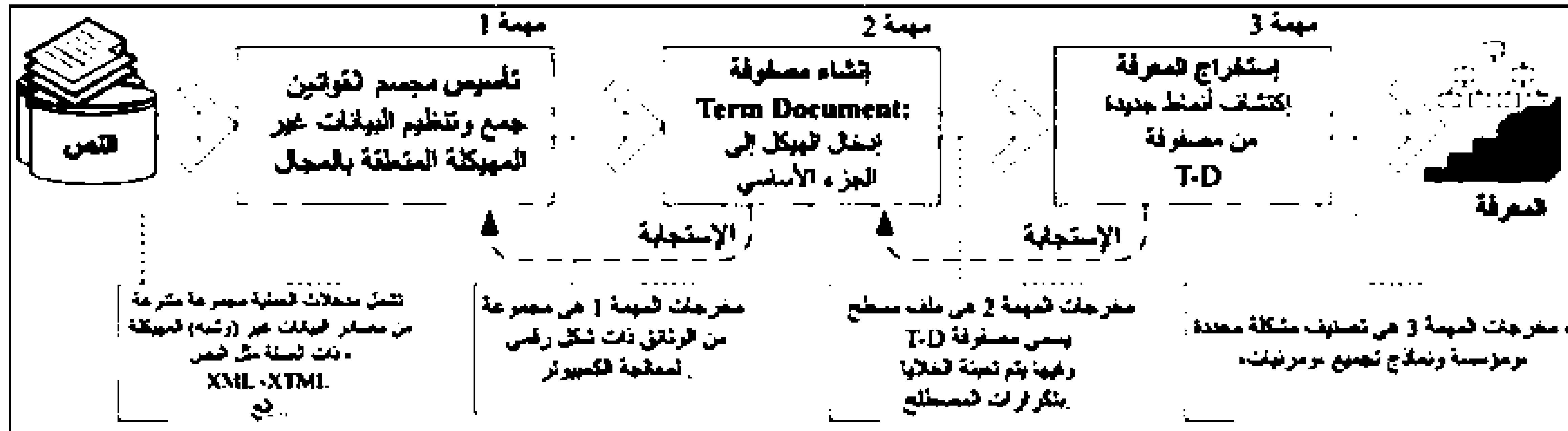
الرغم من أن معظم أجزاء CRISP-DM قابلة للتطبيق أيضًا في مشروعات التنقيب في النص، فإن النموذج المحدد لعملية التنقيب في النص سيشمل أنشطة أكثر تفصيلًا لمعالجة البيانات. ويصور الشكل (٥-٥) مخططًا سياقيًّا عالي المستوى لعملية التنقيب في النص النموذجية (Delen & Crossland, 2008)، حيث يعرض الرسم التخطيطي لهذا السياق نطاق العملية، مع التأكيد على واجهاتها مع البيئة الأكبر. وأما في جوهره، فإنه يرسم حدودًا حول العملية المحددة لتحديد ما يتم إدراجه في (وما يتم استبعاده من) عملية استخراج النص بشكل واضح.



شكل ٥-٥: مخطط السياق لعملية التنقيب في النص

وكما يشير الرسم البياني للسياق، فإن المدخلات (الاتصال الداخلي إلى الحافة اليسرى من المربع) في عملية اكتشاف المعرفة المستندة إلى النص هي البيانات غير المهيكلة وكذلك المهيكلة التي تجمع وتخزن للعملية. والناتج (الامتداد الخارجي من الحافة اليمنى للمربع) من العملية هو المعرفة المحددة للسياق والتي يمكن استخدامها لصنع القرار. وتتضمن عناصر التحكم، التي تسمى أيضًا القيود (الاتصال الداخلي إلى الحافة العلوية للمربع)، العملية الخاصة بالحدود من البرامج والأجهزة، ومشكلات الخصوصية، والصعوبات المتعلقة بمعالجة النص المقدم بلغة طبيعية. أما الآليات فتشمل (الاتصال الداخلي على الحافة السفلية للمربع) الخاصة بالعملية على التقنيات المناسبة وأدوات البرامج وخبرات المجال. والغرض الأساسي من التنقيب في النص (ضمن سياق اكتشاف المعرفة) هو معالجة البيانات غير النصية (مع البيانات المنظمة، إذا كانت ذات صلة بالمشكلة التي يتم تناولها والمتاحة) لاستخراج نماذج ذات معنى وقابلة للتطبيق من أجل اتخاذ قرارات أفضل.

ويمكن تقسـيم عملية التنقيب في النص ذات المستوى العالي جدًا إلى ثلاث مهام متتالية، لكل منها مدخلات محددة لإنتاج نواتج معينة (انظر الشـكل ٥-٦). فإذا لم يوافق ناتج المهمة ما هو متوقع، لأي سبب، فلا بد من الرجوع لتنفيذ المهمة السابقة.



شكل ٥-٦: عملية التنقيب في النص ثلاثية الخطوات والمهام

المهمة ١: تأسيس الجزء الأساسي:

ويتمثل الغرض الرئيس من نشاط المهمة الأولى في جمع كل الوثائق المتعلقة بالسياق (مجال الاهتمام) الذي تتم دراسته. وقد تتضمن هذه المجموعة الوثائق النصية وملفات XML ورسائل البريد الإلكـتروني وصفحات الويب والملاحظات القصيرة. بالإضافـة إلى البيانات النصية المتوفرة بسـهولة، ويمكن أيضًا تسـجيل التسـجيلات الصوتية باسـتخدام خوارزميات التعرف على الكلام وجعلها جزءًا من مجموعة النصوص.

وبمجرد تجميعها، يتم تحويل الوثائق النصية وتنظيمها بطريقة تجعلها كلها في نفس الشـكل التمثيلي (مثل الملفات النصية ASCII) لمعالجة الحاسب. ويمكن أن تكون عملية تنظيم الوثائق بسـيطة مثل مجموعة من مقتطفات النصوص الرقمية المخزنة في مجلد ملف، أو يمكن أن تكون مجموعة من الروابط لمجموعة من صفحات الويب في مجال معين. كما يمكن أن تقبل العديد من أدوات برامـج التنقيب في النص المتاحة تجاريًا كمدخلات وتحويلها إلى ملف مسـطح للمعالجة. وبدلًا من ذلك، فإنه من الممكن إعداد الملف المسـطح خارج برنامج اسـتخراج النص ثم تقديمه كمدخل إلى تطبيق استخراج النص.

المهمة ٢: تأسيس مصفوفة Term-Document:

وفي هذه المهمة، يتم استخدام الوثائق الرقمية والمنظمة (المجموعة) لإنشاء مصفوفة (Term-Document (TDM. حيـث تمثل الصفوف في TDM الوثائق، أمـا الأعمدة فتمثل المصطلحات.

وتتميز العلاقات بين الشروط والوثائق بمؤشرات (أي: مقياس علائقي يمكن أن يكون بسـيطًا مثل عدد مرات حدوث المصطلح في الوثائق ذات الصلة). ويُعد الشكل (٥-٧) مثالًا نموذجيًا على TDM.

المصطلحات / الوثائق	مخاطر الإستثمار	إدارة المشروع	هندسة البرمجيات	التطوير	SAP		
وثيقة 1	1			1			
وثيقة 2		1					
وثيقة 3			3		1		
وثيقة 4		1					
وثيقة 5			2	1			
وثيقة 6	1			1			
...							

شكل ٥-٧: المصفوفة البسيطة Term-Document

ويتمثـل الهدف منها في تحويل قائمة الوثائق المهيكلـة (المجموعة) إلى TDM حيث يتم تعبئة الخلايـا بالمـؤشرات الأكثر ملاءمة. أما الافتراض فهـو أنه يمكن تمثيل جوهر الوثيقـة بقائمة وتكرار المصطلحـات المسـتخدمة في تلك الوثيقة. ومع ذلك، فهل جميـع المصطلحات مهمة عند توصيف الوثائـق؟ والواضـح أن الإجابة هي «لا». فبعـض المصطلحات، مثل: المقالات والأفعال المسـاعدة والمصطلحات المستخدمة في جميع الوثائق في المجموعة، ليس لها أي اختلاف في القوة، وبالتالي يجب استبعادها من عملية الفهرسة. وتُعد هذه القائمة من المصطلحات، والتي يُطلق عليها عادةً مصطلح "شروط التوقف" أو "كلمات التوقف"، خاصة بمجال الدراسة ويجب تحديدها بواسطة خبراء المجال. ومـن ناحية أخرى، فإن المرء قد يختار مجموعة من المصطلحات المحددة سـلفًا والتي بموجبها يتم فهرسة الوثائق (وتسمى هذه القائمة من المصطلحات بشكل ملائم مصطلحات أو قاموس). بالإضافة إلى ذلك، فإنه من الممكن أيضًا توفير المرادفات (أزواج المصطلحات التي سـيتم التعامل معها بنفس الطريقة) والعبارات المحددة (مثل "برج إيفل") بحيث تكون مدخلات الفهرس أكثر دقة.

ومـن عمليات الترشـيح الأخرى التي ينبغي إجراؤها لإنشـاء هذه المؤشرات بدقة، ما يشـير إلى إرجـاع الكلـمات إلى جذورهـا بحيث يتم، على سـبيل المثال تحديد أشـكال نحوية مختلفة أو انحرافات للفعل وفهرسـتها على أنها الكلمة نفسـها. فعلى سـبيل المثال، سوف يضمن النهج التشغيلي أن يكون التشكيل والتصميم معروفًا على غرار نموذج الكلمة.

ويتضمـــن الجيل الأول من TDM جميع المصطلحات الفريدة التـــي تم تحديدها في المجموعة (كأعمدة لها)، باســتثناء تلك الموجــودة في قائمة التوقف؛ بالإضافة إلى جميــع الوثائق (كصفوفها)؛ وعدد التكرارات لكل مصطلح لكل وثيقة (كقيم الخلية الخاصة بها). وكما هو شائع، فإن المجموعة إذا تضمنــت عددًا كبيرًا من الوثائق، فهناك فرصـــة جيدة جدًا لأن يتضمن نموذج TDM عددًا كبيرًا جــدًا من المصطلحات. وقد تســتغرق معالجة مثل هذه المصفوفة الكبــيرة وقتًا طويلًا، والأهم من ذلك، أنها قد تؤدي إلى استخراج أنماط غير دقيقة. وعند هذه النقطة يتعين على المرء أن يقرر ما يلي: (١) ما هو أفضل تمثيل للمؤشرات؟ و(٢) كيف يمكننا تقليل أبعاد هذه المصفوفة إلى حجم معقول؟

تمثيل المؤشرات: بمجرد أن تتم فهرسة وثائق الإدخال وتحويل ترددات الكلمة الأوَّلية (حسب الوثيقة)، فإنه يمكن إجراء عدد من التحويلات الإضافية لتلخيص وتجميع المعلومات المستخرجة. وتعكــس ترددات المصادر الخام عمومًا مدى بروز أو أهمية الكلمة في كل وثيقة. وبوجهٍ خاص، فــإن الكلمات ذات التكرار الأكــبر في الوثيقة هي واصفات أفضل لمحتويــات تلك الوثيقة. ومع ذلك، فمن غير المعقول أن نفترض أن الكلمة تعد نفســها مناسبة مع أهميتها كواصفات للوثائق. فمثــلًا، إذا حدثت كلمة مرة واحــدة في الوثيقة A، في حين أنها في الوثيقة B حدثت ثلاث مرات، فليس من المعقول بالضرورة أن نستنتج أن هذه الكلمة هي ثلاثة أضعاف أهمية واصف الوثيقة B مقارنة بالوثيقة A. وللحصول على نظام TDM أكثر اتســاقًا لإجراء مزيد من التحليل، فيجب أولًا تطبيــع هذه المؤشرات الأوَّلية. وعلى عكــس إظهار التعداد الفعلي للتردد، فإنه يمكن تطبيع التمثيل العددي بين المصطلحات والوثائق باستخدام عدد من الطرق البديلة، مثل ترددات السجل والترددات الثنائية وترددات الوثائق العكسية، وغيرها.

تقليل أبعاد المصفوفة: نظرًا لأن نظام TDM غالبًا ما يكون كبيرًا جدًا ومتناثرًا إلى حد ما (معظم الخلايا مملوءة بالأصفار)، فإن الســؤال الآخر المهم هو «كيف يمكننا تقليل أبعاد هذه المصفوفة إلى حجم معقول؟» حيث تتوفر العديد من الخيارات المتاحة لإدارة حجم المصفوفة.

- يمر خبير المجال بقائمة المصطلحات ويزيل تلك التي لا معنى لها لسياق الدراسة (عملية يدوية تتطلب عمالة كثيرة).
- إلغاء الشروط مع عدد قليل جدًا من الظروف في عدد قليل جدًا من الوثائق.
- تحويل المصفوفة باستخدام انحلال القيمة المفردة SVD.

ويؤدي انحلال القيمة المفردة (SVD)، والذي يرتبط ارتباطًا وثيقًا بتحليل المكونات الأساسية، إلى تقليــل الأبعــاد الإجمالية لمصفوفة الإدخــال (عدد وثائق الإدخال حســب عدد المصطلحات

المستخرجة) إلى مساحة أصغر، حيث يمثل كل بُعد تالي أكبر درجة ممكنة من التقلبية (بين الدرجات والوثائق) (Manning & Schutze, 1999). ومن الناحية المثالية، فإن المحلل قد يحدد البعدين أو الثلاثة أبعاد الأبرز التي تمثل معظم التباين (الاختلافات) بين الكلمات والوثائق، وبالتالي تحديد المساحة الدلالية الكامنة التي تنظم الكلمات والوثائق في التحليل. وبمجرد تحديد هذه الأبعاد، يتم استخراج "المعنى" الأساسي لما هو موجود (تمت مناقشته أو وصفه) في الوثائق.

المهمة ٣: استخراج المعرفة:

حيث يتم استخراج أنماط جديدة في سياق المشكلة المحددة التي يتم تناولها، باستخدام نموذج TDM المنظم جيدًا، والذي يمكن زيادته مع عناصر بيانات منظمة أخرى. أما الفئات الرئيسة لطرق استخلاص المعرفة فهي التصنيف، والتجميع، والاقتران، وتحليل الاتجاهات. وفيما يلي وصف موجز لهذه الطرق.

التصنيف: ويمكن القول أن أكثر موضوعات اكتشاف المعرفة شيوعًا في تحليل مصادر البيانات المعقدة هو تصنيف أشياء معينة. وتتمثل المهمة في تصنيف البيانات المعطاه إلى مجموعة من الفئات المحددة مسبقًا. كما أنه ينطبق على مجال التنقيب في النص، حيث تُعرف المهمة باسم تصنيف النص. فبالنسبة لمجموعة معينة من الفئات (الموضوعات أو النقاط أو المفاهيم) ومجموعة من الوثائق النصية، فإن الهدف هو العثور على الموضوع الصحيح (الموضوع أو المفهوم subject or concept) لكل وثيقة باستخدام النماذج التي تم تطويرها مع مجموعة بيانات التدريب التي تتضمن كلاً من الوثائق وفئات الوثائق الفعلية. واليوم يتم تطبيق تصنيف النص التلقائي في مجموعة متنوعة من السياقات، بما في ذلك فهرسة النص التلقائي أو شبه التلقائي (التفاعلي)، وتصفية الرسائل غير المرغوب فيها، بالإضافة إلى تصنيف صفحات الويب ضمن كتالوجات هرمية، وتوليد تلقائي للبيانات الوصفية، واكتشاف النوع، فضلًا عن العديد من الأنواع الأخرى.

وأما النهجان الرئيسان لتصنيف النص، فهما هندسة المعرفة وتعلم الآلة (Feldman & Sanger, 2007). وباستخدام منهج هندسة المعرفة يتم ترميز معرفة الخبير حول الفئات في النظام إما بصورة عرضية أو في شكل قواعد تصنيف إجرائية. أما مع منهج تعلم الآلة، فإن العملية الاستقرائية العامة تبني مصنفًا من خلال التعلم من مجموعة من الامثلة المُعاد تصنيفها. ومع زيادة عدد الوثائق بمعدل أُسي، ومع ازدياد صعوبة معرفة خبراء المعرفة، فإن الاتجاه الأكثر انتشارًا بين الاثنين هو الاتجاه نحو نهج تعلم الآلة.

التجميــع: وهــو عملية غير خاضعة لــلإشراف؛ حيث يتم تصنيف الكائنــات إلى مجموعات «طبيعية» تسمى المجموعات، بالمقارنة مع التصنيف الذي يتم فيه استخدام مجموعة من الأمثلة التدريبية المصنفة مســبقًا لتطوير نموذج يعتمد على الســمات الوصفيــة للفئات لتصنيف مثال جديد غير مذكور. أما في التجميع، فإن المشكلة تتمثل في تجميع مجموعة غير معلمة من الأشياء (مثل: الوثائق، والزبون، والتعليقات، وصفحات الويب) في مجموعات ذات مغزى دون أي معرفة مسبقة.

ويُعــد التجميع مفيدًا في نطاق واســع مــن التطبيقات، بدءًا من اســترجاع الوثائق إلى تمكين عمليــات بحث أفضل في محتــوى الويب. وفي الواقع، فإن أحد التطبيقــات البارزة للتجميع هو التحليل والإبحار في مجموعات نصية كبيرة جدًا، مثل صفحات الويب. والافتراض الأساسي هو أن الوثائق ذات الصلة تميل إلى أن تكون أكثر تشــابهًا مع بعضها البعض من غير التي ليس لها صلة بالموضوع. فإذا كان هذا الافتراض قائمًا، فإن تجميع الوثائق اعتمادًا على تشــابه محتواها يُحسِّن فعالية البحث (Feldman & Sanger, 2007):

- **استرجاع بحث تم تحسينه:** ولأن التجميع يستند إلى تشابه عام بدلًا من وجود مصطلح واحد، فإنه من الممكن أن يؤدي إلى تحســين اســترجاع بحث معتمد على الاســتعلام بطريقة تطابق استعلام وثيقة يمكن استرجاع المجموعة الكاملة منها.
- **تحسين دقة البحث:** حيث يمكن للتجميع أيضًا أن يحسن دقة البحث. فمع نمو عدد الوثائق في مجموعة، يصبح من الصعب اســتعراض قائمة الوثائق المطابقة. ويمكن أن يســاعد التجميع على تجميــع الوثائــق في عدد من المجموعات الأصغر حجمًا من الوثائق ذات الصلة، وطلبها حســب الصلة وعرض الوثائق من المجموعة (أو المجموعات) الأكثر صلة.

وأمــا الطريقتــان الأكثر شــيوعًا للتجميع فهما التجميــع المبعثر / المجمــع والتجميع الخاص بالاستعلام:

- **التجميع المبعثر:** حيث يُستخدم التجميع كطريقة لتصفح الوثائق وذلك لتحسين كفاءة التصفح البــشري للوثائــق عندما لا يمكن صياغة اســتعلام بحث معين، بمعنى أن هــذه الطريقة تقوم بشــكل ديناميكي بإنشــاء جدول محتويات لمجموعة وتقوم بتكيفها وتعديلها استجابةً لاختيار المستخدم.
- **التجميع الخاص بالاســتعلام Query-specific clustering:** وتســتخدم هذه الطريقة نهجًا تجميعيًــا للتسلســل الهرمــي حيث تظهر أكــثر الوســائل ذات الصلة بالاســتعلام المطروح في

مجموعــات ضيقة صغــيرة متداخلة في مجموعات أكــبر تحتوي على وثائق أقل تشــابهًا، مما يؤدي إلى وجود مجموعة من مســتويات الملاءمة بين الوثائق. وتؤدي هذه الطريقة أداءً جيدًا لمجموعات الوثائق ذات الأحجام الكبيرة واقعيًّا.

الاقــتران: وقد تم تقديم تعريف رســمي ووصف مفصل للاقتران في الفصل الخاص بالتنقيب في البيانــات (الفصل الرابع). يُعد تعليم الارتباط أو قواعد الارتباط في مجال التنقيب في البيانات أســلوبًا شــائعًا وبحثًا جيدًا لاكتشــاف علاقات مثــيرة للاهتمام بين المتغــيرات في قواعد البيانات الكبيرة. وتتمثل الفكرة الرئيسة في إنشاء قواعد الارتباط (أو حل مشكلات سلة السوق) في تحديد المجموعات المتكررة التي تتماشى معًا.

وفي التنقيب في النص، يشير الارتباط تحديدًا إلى العلاقات المباشرة بين المفاهيم (المصطلحات) أو مجموعات المفاهيم. ويمكن تحديد مفهوم مجموعة قواعد القاعدة A+C المتعلقة بمجموعتَي المفهوم المتكررتين A وC بواسطة المقياسين الأساسيين للدعم والثقة. وفي هذه الحالة، فإن الثقة هي النسبة المئوية للوثائق التي تتضمن جميع المفاهيم في C داخل نفس المجموعة الفرعية من تلــك الوثائق التي تتضمن جميع المفاهيم في A. أما الدعم فهو النســبة (أو العدد) من الوثائق التــي تتضمن جميع المفاهيم في A. وعلى ســبيل المثال، فقد يظهــر في مجموعة الوثائق مفهوم "فشل تنفيذ البرامج" غالبًا مقترنًا مع "تخطيط موارد المؤسسة" و"إدارة علاقات العملاء" بدرجة دعم معنوي (٤٪) ومستوى ثقة (٥٥٪)، وهذا يعني أن ٤٪ من الوثائق كانت تشتمل على جميع المفاهيم الثلاثة ممثلة في نفس الوثيقة، والوثائق التي شملت "فشل في تنفيذ البرامج"، ٥٥٪ منها شملت أيضًا "تخطيط موارد المؤسسات" و"إدارة علاقات العملاء".

وقد تم اســتخدام التنقيب في النص مــع قواعد الاقتران لتحليل الأدبيات المنشــورة (الأخبار والمقالات الأكاديمية المنشورة على الويب) لرسم تَفشِّي مرض أنفلونزا الطيور وتطوره (Mahgoub وآخرون، ٢٠٠٨). كانت الفكرة هي تحديد الارتباط التلقائي بين المناطق الجغرافية، والانتشار عبر الأنواع، والتدابير المضادة (العلاجات).

تحليلات الاتجاه: تســتند الأســاليب الحديثة لتحليلات الاتجاه في مجال التنقيب في النص على فكــرة أن أنــواع مختلفة مــن توزيعات المفهوم هي وظائــف مجموعات الوثائــق، وهذا يعني أن مجموعات مختلفة تؤدي إلى توزيعات مختلفة للمفهوم لنفس مجموعة المفاهيم. ولذلك؛ فإنه من الممكن مقارنة التوزيعين اللذين يتطابقان بشكل ما إلا أنهما من توزيعات فرعية مختلفة. ويتمثل أحد الاتجاهات الملحوظة لهذا النوع من التحليل في وجود مجموعتين من نفس المصدر (على سبيل

المثـال من نفس مجموعة الدوريات الأكاديمية) ولكن مـن نقاط زمنية مختلفة. وقد طبق كل من Delen وCrossland (٢٠٠٨) تحليـل الاتجـاه على عدد كبير من المقالات الأكاديمية (المنشـورة في المجلات الأكاديمية الثلاث الأعلى تصنيفًا) لتحديد تطور المفاهيم الأساسية في مجال نظم المعلومات.

وكما هو موضح في هذا القسم، فإن هناك العديد من الطرق المتاحة للتنقيب في النص. وتصف الحالة العملية (٥-٥) استخدام عدد من التقنيات المختلفة في تحليل مجموعة كبيرة من الأدبيات.

حالة عملية ٥-٥

مسح بحوث الأدبيات بالتنقيب في النص

يواجه الباحثون الذين يجرون عمليات بحث واستعراض للأدبيات ذات الصلة مهمة معقدة وضخمة على نحو متزايد. وقد كان من المهم دائمًا العمل بجد لجمع المعلومات الحالية من الأدبيات وتنظيمها وتحليلها واستيعابها، من أجل توسيع نطاق المعرفة ذات الصلة، لا سـيما من خلال الانضباط المنزلي. ومع تزايد وفـرة الأبحاث الهامة المحتملة التـي يتم الإبلاغ عنهـا في المجالات ذات الصلة، وحتى في المجـالات التي تعتبر تقليديًا حقـولًا غير مرتبطة بالدراسـة، فإن مهمة الباحث تصبح أكـثر صعوبة، إذا كانت هناك حاجة إلى عمل شامل.

وفي تيـارات البحـث الجديدة، قد تكـون مهمة الباحث أكثر ملـلًا وتعقيدًا. ففي أحسـن الأحوال، قد يكون مـن الصعب محاولة العثور على الأعـمال ذات الصلة التي ذكرها الآخرون، بل وربما يكون شـبه مسـتحيل إلى حدٍ كبير إذا كانت هناك حاجة إلى مراجعات تقليدية، للأبحاث المنشـورة. فحتى مع وجود مجموعة من طلاب الدراسات العليا المتفانين أو الزملاء المسـاعدين، فإن محاولة تغطية جميع الأعمال المنشورة ذات الصلة المحتملة هي إشكالية.

وهناك العديد من المؤتمرات العلمية التي تُعقد كل عام، بالإضافة إلى توسـيع نطاق المعرفـة بالتركيز الحالي على المؤتمر، حيث يرغب المنظمون في كثير من الأحيان في تقديم minitracks وورش عمـل إضافيـة. وفي العديـد من الحالات، تهـدف هذه الأحداث الإضافيـة إلى تعريف الحضور على تيارات بحث مهمة في مجالات الدراسـة ذات الصلة ومحاولـة تحديد "الشيء الكبير التالي" من حيث الاهتمامات البحثية والتركيز. وغالبًا ما يكـون تحديد مواضيع المرشـح المعقول لمثل هذه الثـورات وورش العمل ذاتيًا بدلًا من اشتقاق موضوعي من الأبحاث الحالية والناشئة.

وفي دراسة حديثة، اقترح كل من Delen وCrosslan (٢٠٠٨) طريقة تساعد جهود الباحثين بشكل كبير وتعززها من خلال تمكين التحليل شبه الآلي للكميات الكبيرة من الأبحاث المنشورة من خلال تطبيق التنقيب في النص. وباستخدام المكتبات الرقمية القياسية ومحركات بحث النشر على الإنترنت، قام المؤلفون بتنزيل وجمع جميع المقالات المتاحة للمجلات الثلاث الرئيسة في مجال نظم المعلومات الإدارية: MISQ ربع السنوي (MISQ)، وبحوث نظم المعلومات (ISR)، ومجلة نظم معلومات الإدارة (JMIS). وللحفاظ على الفترة الزمنية نفسها لجميع المجلات الثلاث (للدراسات الطولية المقارنة المحتملة)، تم استخدام المجلة مع تاريخ البدء الأخير لتوفر النشر الرقمي كوقت بدء هذه الدراسة (أي أن مقالات JMIS متوفرة رقميًا منذ ١٩٩٤)، بالنسبة لكل مقالة، حيث قاموا باستخلاص العنوان، والملخص، وقائمة المؤلفين، والكلمات المفتاحية المنشورة، والحجم، ورقم الإصدار، وسنة النشر، ثم قاموا بتنزيل كل بيانات المقالة في ملف قاعدة بيانات بسيط. كما تم تضمينه في مجموعة البيانات المدمجة التي لها مجال مشترك، والذي يحدد نوع المجلة لكل مادة من أجل تحليل تمييزي محتمل. كما تم حذف الملاحظات التحريرية والملاحظات البحثية والنظرات العامة التنفيذية من المجموعة. يوضح الجدول التالي كيفية عرض البيانات بتنسيق جدولي.

وقد اختاروا في مرحلة التحليل استخدام ملخص للمقالة فقط كمصدر لاستخراج المعلومات. كما اختاروا عدم تضمين الكلمات المفتاحية المدرجة في المنشورات لسببين رئيسين: (١) في ظل الظروف العادية، فإن الملخص سيشمل بالفعل الكلمات المفتاحية المدرجة، وبالتالي فإن إدراج الكلمات المفتاحية المدرجة للتحليل سيعني تكرار نفس المعلومات ويحتمل أن يمنحهم وزنًا غير مستحق، (٢) قد تكون الكلمات المفتاحية المدرجة عبارة عن كلمات يرغب المؤلفون في أن ترتبط مقالتهم بها (على عكس ما هو موجود فعلًا في المقالة)، وبالتالي قد يكون من المحتمل تقديم تحيز غير قابل للتحليل لتحليل المحتوى.

ID	العام	مجلة	نبذة مختصرة
PID001	2005	MISQ	الحاجة إلى ابتكار قيم مستمرة هي دفع سلاسل التوريد للتطور من ....
PID002	1999	ISR	على الرغم من أن الكثير من الأفكار المعاصرة تعتبر تقنية المعلومات المتقدمة.
PID003	2001	JMIS	عندما يواجه منتجو البضائع (أو الخدمات) وضعًا والذي ...

ID	العام	مجلة	نبذة مختصرة
PID004	1995	ISR	أصبحت أهمية الحفاظ على الذاكرة التنظيمية مهمة لـ ...
PID005	1994	ISR	البحث الوارد هنا هو تكيف لنموذج تم تطويره لقياس ...
PID006	1995	MISQ	تقيم هذه الدراسة مدى القيمة المضافة للعملاء من ...
PID007	2003	MISQ	هذه الورقة هي تقارير لنتائج (-) من دراسة ميدانية لستة فِرَق للمشروع الطبي ...
PID008	1999	JMIS	بدأ الباحثون والمديرون يدركون أن المزايا الكاملة ....
PID009	2000	JMIS	تقنيات التجارة عبر الإنترنت قد خفَّضت إلى حدٍ كبير تكاليف البائعين.
PID0010	1997	ISR	نظرية التركيب التكيفية (AST) أصبحت بسرعة نظرية مؤثرة.
PID0011	1995	JMIS	يظهر البحث أن نظم دعم المجموعة (GSS) قد زادت بشكل كبير.
PID0012	2000	MISQ	يطالب سير الأعمال بشكل متزايد بأن تلعب تقنية المعلومات دور الأعمال التجارية.
PID0013	2001	ISR	من المعتقد على نطاق واسع أن التوافق بين إستراتيجية العمل وإستراتيجية IS.
PID0014	1999	JMIS	تم توضيح إطار يتضمن تخطيط أهداف تقنية المعلومات وتحديدها.
PID0015	1999	JMIS	الأهمية المتزايدة باستمرار لتقنية المعلومات (IT) تتطلب ...
PID0016	1994	MISQ	تحديد أفضل طريقة لتنظيم وظائف IS داخل المؤسسة.
PID0017	1996	ISR	أسباب ردود الفعل المختلطة لتداول اليوم الإلكتروني غير التبادلي.
PID0018	1996	JMIS	آثار الأداء لاستثمارات تقنية المعلومات في المنظمة.

ID	العام	مجلة	نبذة مختصرة
PID0019	1997	JMIS	عدم الكشف عن الهوية هو مفهوم أساسي في بحوث نظم دعم المجموعة (GSS).
PID0020	2002	ISR	على الرغم من أن التجارة الإلكترونية (EC) قد خلقت فرصًا جديدة للعمل.
PID0021	2005	JMIS	فهم الاعتماد الناجح لتقنية المعلومات إلى حدٍ كبير.
PID0022	2005	MISQ	نظم تخطيط موارد المؤسسة (ERP) وغيرها من المعلومات المعقدة.
PID0023	1994	JMIS	نظم إدارة نموذج دعم نماذج في مختلف مراحل النموذج.
PID0024	1995	ISR	بينما يتم التعرف على تدريب الحاسب على نطاق واسع باعتباره الحاسب الأساسي.

وقد كانت الدراسة الاستكشافية الأولى هي النظر في المنظور الطولي للمجلات الثلاث (أي تطوير موضوعات البحث بمرور الوقت). ومن أجل إجراء دراسة طولية، قسموا فترة الـ ١٢ سنة (من ١٩٩٤ إلى ٢٠٠٥) إلى أربع فترات بحيث تكون كل فترة عبارة عن ثلاث سنوات لكل من المجلات الثلاث. وقد أدى هذا الإطار إلى ١٢ تجربة للتنقيب في النص مع ١٢ مجموعة من البيانات الحصرية المتبادلة. عند هذه النقطة، وقد استخدمت كل مجموعة من مجموعات البيانات الاثنتي عشرة التنقيب في النص لاستخراج أكثر العبارات وصفية من هذه المجموعات من المقالات التي تمثلها ملخصاتها، وتم جدولة النتائج وفحصها للتغييرات الزمنية المتغيرة في الشروط المنشورة في هذه المجلات الثلاث.

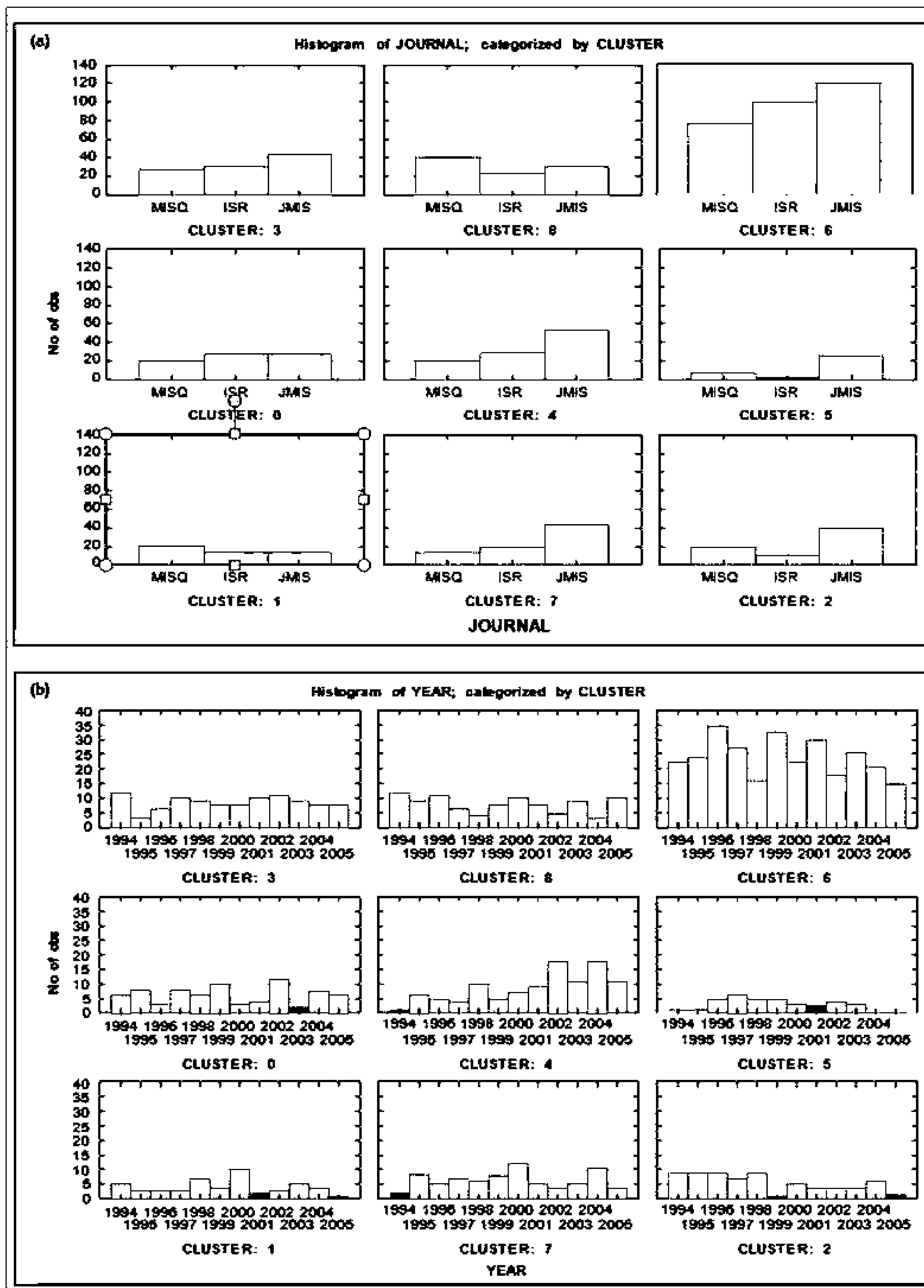
وكاستكشاف ثانٍ، باستخدام مجموعة البيانات الكاملة (بما في ذلك المجلات الثلاث وجميع الفترات الأربع)، قاموا بإجراء تحليل للتجميع. ويمكن القول أن التجميع هو أكثر تقنيات معالجة النصوص استخدامًا. وقد تم استخدام التجميع في هذه الدراسة لتحديد المجموعات الطبيعية للمقالات (عن طريق وضعها في مجموعات منفصلة) ثم سرد أكثر العبارات وصفية التي ميزت تلك المجموعات. كما استخدموا SVD لتقليل الأبعاد من مصفوفة الوثائق من قبل، ومن ثم خوارزمية تعظيم التوقعات لإنشاء التجمعات. وقد أجريت العديد من التجارب لتحديد العدد الأمثل للمجموعات، والتي

تبين أنها تسعة. وبعد بناء التجمعات التسعة، قاموا بتحليل محتوى تلك المجموعات من منظورين: (١) تمثيل نوع المجلة (انظر الشكل ٥-٨أ)، (٢) تمثيل الوقت (الشكل ٥-٨ب). وكانت الفكرة هي استكشاف الاختلافات المحتملة والقواسم المشتركة بين المجلات الثلاث والتغييرات المحتملة في التركيز على هذه المجموعات؛ أي للإجابة على أسئلة مثل: "هل هناك مجموعات تمثل مواضيع بحث مختلفة خاصة بمجلة واحدة؟" و"هل هناك توصيف زمني متفاوت لهذه المجموعات؟" اكتشفوا وناقشوا عدة أنماط مثيرة للاهتمام باستخدام تمثيل جدولي ورسوم بيانية من النتائج التي توصلوا إليها (ولمزيد من المعلومات، انظر: Delen & Crossland, 2008).

أسئلة للمناقشة:

١- كيف يمكن استخدام التنقيب في النص لتسهيل المهمة التي لا يمكن التغلب عليها لمراجعة الأبحاث المنشورة؟

٢- ماهي النتائج الشائعة لمشروع التنقيب في النص لمجموعة محددة من مقالات الدوريات؟ وهل يمكنك التفكير في النتائج المحتملة الأخرى التي لم يتم ذكرها في هذه الحالة؟



Source: Delen, D., & Crossland, M. (2008). Seeding the survey and analysis of research literature with text mining. Expert Systems with Applications, 34(3), 1707- 1720.

الشكل ٥-٨: أ- توزيع عدد المقالات للمجلات الثلاث على المجموعات التسع

ب- تطوير المجموعات التسع على مر السنين

أسئلة مراجعة على القسم ٥-٥:

١- ما هي الخطوات الرئيسة في عملية التنقيب في النص؟

٢- ما هو سبب اعتدال ترددات الكلمات؟ وما هي الطرق الشائعة لاعتدال ترددات الكلمات؟

٣- ما هو SVD؟ وكيف يتم استخدامه في التنقيب في النص؟

٤- ما هي طرق استخراج المعرفة الأساسية من مجسم القوانين؟

٥-٦ تحليل المشاعر:

نعتبر نحن البشر كائنات اجتماعية بارعة في الاستفادة من مجموعة متنوعة من وسائل التواصل، وكثيرًا ما نتشاور مع منتديات المناقشة المالية قبل اتخاذ قرار الاستثمار؛ فنسأل أصدقاءنا عن آرائهم حول مطعم تم افتتاحه حديثًا أو فيلم تم إنتاجه حديثًا؛ كما نقوم بإجراء عمليات بحث على الإنترنت وقراءة تعليقات المستهلكين وتقارير الخبراء قبل إجراء عملية شراء كبيرة مثل منزل أو سيارة أو جهاز. ونحن نعتمد على آراء الآخرين لاتخاذ قرارات أفضل، خاصة في منطقة لا نملك فيها الكثير من المعرفة أو الخبرة. وبفضل تزايد توافر موارد الإنترنت الغنية بالرأي مثل وسائل التواصل الاجتماعي (مثل Twitter وFacebook) ومواقع المراجعة على الإنترنت والمدونات الشخصية، أصبح الآن من السهل أكثر من أي وقت مضى العثور على آراء الآخرين (الآلاف منهم، كما هو واقع الأمر) على كل شيء من أحدث الأدوات إلى الشخصيات السياسية والعامة. وعلى الرغم من عدم التعبير عن آراء عبر الإنترنت - بسبب الغالبية المتزايدة لعدد قنوات التواصل الاجتماعي وقدراتها - إلا أن الأرقام تتزايد بشكل كبير. ومن الصعب وضع تعريف محدد لكلمة المشاعر، حيث إنها غالبًا ما ترتبط بمصطلحات أخرى أو تختلط معها مثل الاعتقاد، والرؤية، والرأي، بالإضافة إلى القناعة. وتشير المشاعر إلى وجود رأي ثابت يعكس مشاعر المرء (Mejova, 2009). وللمشاعر بعض الخصائص الفريدة التي تميزها عن المفاهيم الأخرى التي قد نرغب في تحديدها في النص. وغالبًا ما نرغب بتصنيف النص حسب الموضوع، والذي قد يتضمن التعامل مع تصنيفات كاملة للموضوعات. ومن ناحية أخرى، فعادةً ما يقوم تصنيف المشاعر بالتعامل مع فئتين (إيجابي مقابل سلبي)، أو نطاق من التناقض (على سبيل المثال، تصنيف النجوم للأفلام)، أو حتى نطاق قوة الرأي (Pang & Lee, 2008). وتمتد هذه الفئات لتشمل العديد من المواضيع والمستخدمين والوثائق. وعلى الرغم من أن التعامل مع فئات قليلة فقط قد يبدو مهمة أسهل من تحليل النص القياسي، إلا أن هذا أبعد ما يكون عن الحقيقة.

وكمجال للبحث، فإن تحليل المشاعر يرتبط ارتباطًا وثيقًا باللسانيات الحاسوبية، معالجة اللغات الطبيعية NLP، والتنقيب في النصوص. ولتحليل المشاعر العديد من الأسماء، وغالبًا ما يشار

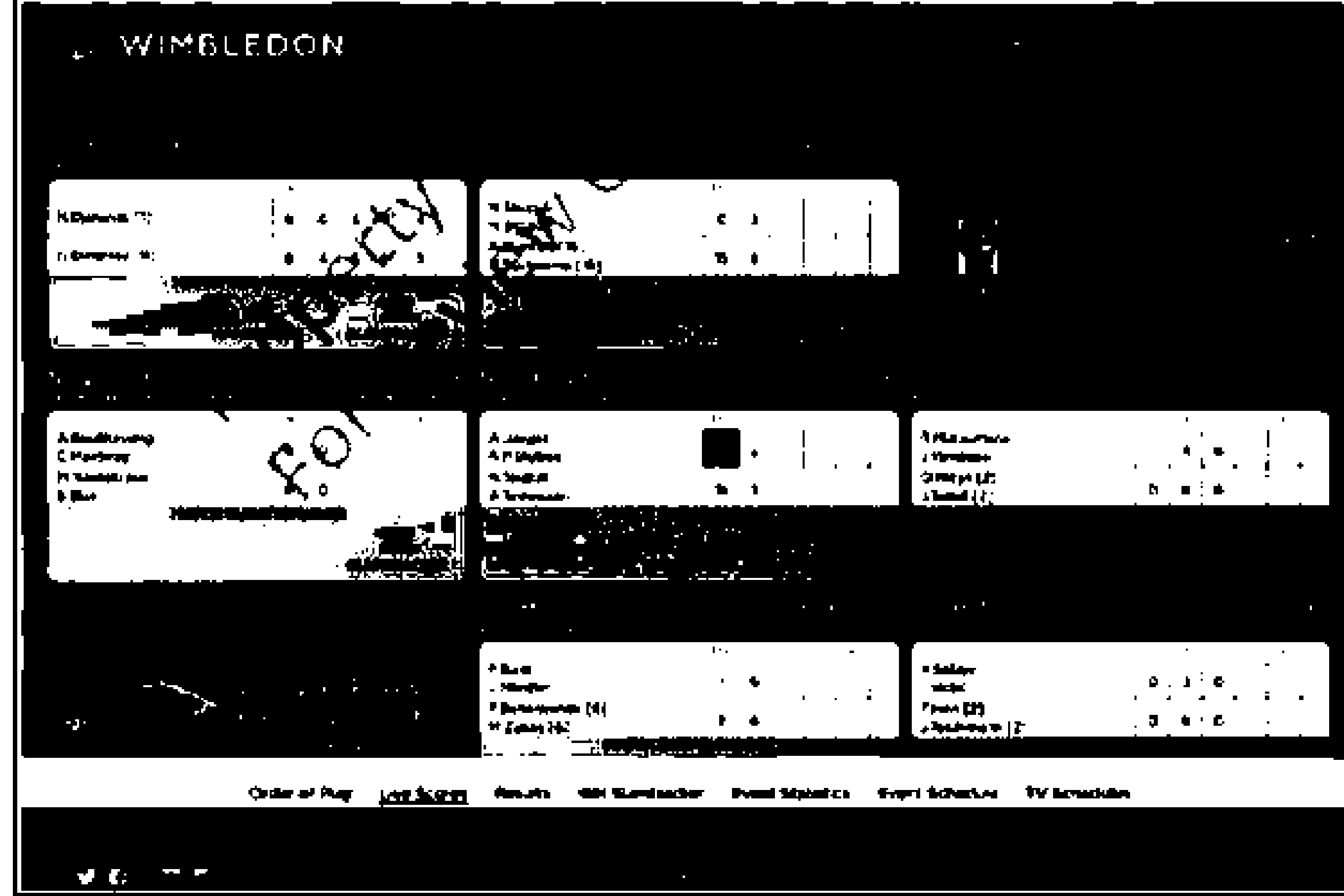
إليه على أنه استخراج الرأي، وتحليل الذاتية، واستخراج التقييم، مع بعض الاتصالات بالحوسبة العاطفية (التعرف على الحاسب والتعبير عن الانفعال). إن الاندفاع المفاجئ في الاهتمام والنشاط في مجال تحليل المشاعر (أي استخراج الرأي)، الذي يتناول الاستخراج التلقائي للآراء والمشاعر والديموقراطية في النص، ينتج فرصًا وتهديدات للشركات والأفراد على حد سواء، ولن يستفيد منها سوى الأشخاص الذين يتقبلونها وينتفعون منها ... وسوف يتم اعتماد كل رأي يتم طرحه على الإنترنت من قِبل فرد أو شركة إلى المنشئ (جيد أو سيئ) كما سيتم استرداده واستخراجه بواسطة الآخرين (غالبًا ما يتم تلقائيًّا بواسطة برامج الحاسب).

ويحاول تحليل المشاعر الإجابة عن السؤال التالي «ما الذي يشعر به الناس تجاه موضوع معين؟» عن طريق البحث في آراء الكثيرين باستخدام مجموعة متنوعة من الأدوات الآلية، وذلك من خلال الجمع بين الباحثين والممارسين في مجالات الأعمال وعلوم الحاسب واللغويات الحاسوبية والتنقيب في البيانات والتنقيب في النصوص بالإضافة إلى علم النفس وحتى علم الاجتماع. ويهدف تحليل المشاعر إلى توسيع نطاق تحليل النصوص القائم على الحقائق التقليدية إلى حدود جديدة لتحقيق نظم المعلومات الموجهة نحو الرأي العام. وفي إطار الأعمال التجارية، خاصة في التسويق وإدارة علاقات العملاء، يسعى تحليل المشاعر إلى الكشف عن الآراء الإيجابية وغيرها تجاه منتجات أو خدمات معينة باستخدام أعداد كبيرة من مصادر البيانات النصية (ملاحظات العملاء في شكل نشرات الويب، وتويتر، والمدونات، إلخ).

ويأتي الشعور الذي يظهر في النص بصفتين مميزتين: كونه صريحًا؛ حيث تعبر الجملة الشخصية مباشرة عن رأي ("إنه يوم رائع")، كما أنه ضمنيًّا؛ حيث يشير النص إلى رأي ("ينفصل المقبض بسهولة كبيرة"). وقد ركز معظم العمل السابق في تحليل المشاعر على النوع الأول من المشاعر لأنه أسهل في التحليل. أما الاتجاهات الحالية، فهي تنفيذ طرق تحليلية للنظر في المشاعر الضمنية والصريحة. إن تناقض المشاعر هي سمة خاصة للنص الذي يركز تحليل المشاعر في المقام الأول عليه، وعادةً ما يتم تقسيمها إلى قسمين: موجبة وسالبة، ولكن يمكن اعتبار الأقطاب أيضًا كنطاق. وسوف يعتري الوثيقة التي بها العديد من البيانات المرتقبة خلط في التناقض الكامل، وهو يختلف عن عدم وجود تناقض مطلقًا (كونها موضوعية؛ Mejova, 2009). إن جمع وتحليل البيانات النصية في الوقت المناسب، والتي قد تأتي من مجموعة متنوعة من المصادر - بدءًا من نسخ مراكز اتصال العملاء إلى منشورات وسائل التواصل الاجتماعي - تعد جزءًا هامًا من قدرات الشركات الاستباقية والتركيز على العملاء في الوقت الحاضر، وفي كثير من الأحيان يتم تصوير هذه التحليلات بشكل فوري للبيانات النصية في لوحات المعلومات سهلة الفهم. وتقدم الحالة العملية (٥-٦) قصة نجاح العميل، حيث يتم استخدام مجموعة من حلول التحليلات جماعيًّا لتعزيز تجربة المشاهدين في بطولة ويمبلدون للتنس.

حالة عملية ٥-٦

إنشاء تجربة رقمية فريدة لرصد اللحظات المهمة في ويمبلدون



وتُعرف ببساطة لملايين من المشجعين باسم «ويمبلدون»، وهي أقدم بطولات التنس الأربع الكبرى، وهي واحدة من أرقى الأحداث الرياضية في العالم، وقد تم تنظيمها من قِبل نادي All England Lawn Tennis Club (AELTC)، والذي يُعد مؤسسة رياضية وثقافية عالمية منذ عام ١٨٧٧.

بطل البطولات:

إن منظمي بطولات ويمبلدون، وAELTCK، لديهم هدف بسيط: فهم يريدون استضافة أفضل بطولات التنس في العالم كل عام - بكل طريقة، وبكل مقياس.

والدافع وراء هذا الالتزام ليس مجرد الكبرياء؛ كما أن لديها أساسًا تجاريًا. فقد تم بناء علامة ويمبلدون (Wimbledon) التجارية على مكانتها الرائدة، وهذا ما يجذب كلًا من المعجبين والشركاء. وتسعى أفضل المنظمات الإعلامية في العالم وأعظم الشركات - بما في ذلك IBM - أن تكون مرتبطة مع ويمبلدون على وجه التحديد بسبب سمعتها للتميز.

ولهذا السبب، يعد الحفاظ على هيبة البطولة من أهم أولويات AELTC، ولكن هناك طريقتان فقط يمكن أن تتحكم فيهما المنظمة مباشرة في كيفية النظر إلى بقية البطولات من قِبل بقية العالم.

الأولى والأكثر أهمية هي: توفير تجربة رائعة للاعبين والصحفيين والمتفرجين المحظوظين بما يكفي لزيارة ومراقبة ملعب التنس. وتتمتع AELTC بخبرة كبيرة في

هذا المجال. ففي عام ١٨٧٧، أقامت منافسة مثيرة ورائعة على مدار أسبوعين في أجواء شاعرية: التنس في الريف الإنجليزي.

والثانية هي: الحضور للبطولة عبر الإنترنت، والذي يتم تسليمه عبر موقع الويب wimbledon.com وتطبيقات الجوال وقنوات التواصل الاجتماعي. وهذا التطور المستمر لهذه المنصات الرقمية هو نتيجة شراكة مدتها ٢٦ عامًا بين AELTC وIBM.

ويوضح Mick Desmond، المدير التجاري والإعلامي في AELTC قائلًا: «عندما تشاهد ويمبلدون على التلفاز، فإنك تشاهده من خلال عدسة المذيع. ونحن نفعل كل ما في وسعنا لمساعدة شركائنا في وسائل التواصل على تقديم أفضل عرض ممكن. ويختلف النظام الرقمي؛ فهو نظامنا الأساسي، حيث يمكننا التحدث مباشرة مع المعجبين لدينا، لذلك فمن المهم أن نقدم لهم أفضل تجربة ممكنة، وإذا أردنا تعزيز علامتنا التجارية، فنحن بحاجة إلى أن يشاهدوا تجربتنا الرقمية كمركز رقم واحد لمتابعة البطولة على الإنترنت".

ولذا؛ فقد استهدفت AELTC جذب ٧٠ مليون زيارة، و٢٠ مليون جهاز فريد، و٨ ملايين متابع خلال أسبوعين من بطولة ٢٠١٥، وكان الأمر متروكًا لشركتي IBM و AELTC لإيجاد طريقة للتوصيل.

نقل تجربة رقمية فريدة:

شرعت IBM وAELTC في إعادة تصميم كاملة للمنصة الرقمية، وذلك باستخدام معرفتهم الوثيقة بجماهير البطولات بهدف تطوير تجربة مصممة خصيصًا لجذب مشجعي التنس من جميع أنحاء العالم والإبقاء عليهم.

تقول Alexandra Willis، رئيس قسم المحتوى الرقمي في AELTC: «لقد أدركنا أنه على الرغم من الأهمية المتزايدة للهاتف المحمول، فإن ٨٠٪ من زوارنا يستخدمون أجهزة الحاسب المكتبية للوصول إلى موقعنا على الويب، وكان التحدي الذي واجهنا لعام ٢٠١٥ هو كيفية تحديث خصائصنا الرقمية للتكيف مع عالم متنقل، مع توفير أفضل تجربة ممكنة لسطح المكتب على الإطلاق. وقد أردنا لموقعنا الجديد أن يستفيد إلى أقصى حد من تلك الشاشة الكبيرة، ويمنح مستخدمي سطح المكتب أغنى تجربة ممكنة فيما يتعلق بالمشاهد المرئية ومحتوى الفيديو عالي الوضوح - بينما يتفاعلون أيضًا ويتكيفون بسهولة مع أشكال الأجهزة اللوحية أو الهواتف المحمولة الأصغر».

وتضيف: «ثانيًا، ركزنا تركيزًا عاليًا على وضع محتوى في السياق، أي تكامل المقالات مع الصور ومقاطع الفيديو والإحصاءات ومقتطفات المعلومات ذات الصلة، وتبسيط التنقل بحيث يمكن للمستخدمين الانتقال بسلاسة إلى المحتوى الأكثر أهمية لهم».

وبالنسبة لأجهزة المحمول، فقد أدرك الفريق أن إتاحة اتصالات 4G بسعة تحميلية عالية يعني أن موقع الويب للمحمول سيصبح أكثر شعبية من أي وقت مضى، كما سيضمن سهولة الوصول إلى جميع محتوى الوسائط الغنية. وفي الوقت نفسه، تم تحسين تطبيقات المحمول في البطولات بإشعارات لنتائج المباريات في وقت حدوثها، بل ويمكنها أيضًا تقديم التحية للزوار أثناء مرورهم عبر المحطات في الطريق إلى الملاعب.

كما قام الفريق ببناء مجموعة خاصة من المواقع الإلكترونية لأهم جماهير للعبة التنس، وهي اللاعبين أنفسهم؛ حيث قاموا - باستخدام تقنية ®IBM ®Bluemix- ببناء تطبيق ويب آمن يوفر للاعبين عرضًا شخصيًا لحجوزاتهم في الملعب، وتنقلاتهم، وأوقات اللعب، بالإضافة إلى مساعدتهم في مراجعة أدائهم بالوصول إلى إحصائيات كل مباراة لعبوها.

تحويل البيانات إلى صورة مرئية - والصورة المرئية إلى قصة:

ولتزويد منصاتها الرقمية بالمحتوى الأكثر إقناعًا، فقد استفاد الفريق من ميزة فريدة، وهي وصوله إلى البيانات الخاصة بكل تسديدة في كل مباراة خلال البطولة في وقت حدوثها. وعلى مدار أسبوعين من بطولة ويمبلدون، قام ٤٨ خبيرًا تحكيميًا بجمع ما يقرب من ٣,٤ مليون نقطة بيانات، وتتبع نوع اللقطة، والإستراتيجيات، ونتائج كل نقطة.

ويتم جمع هذه البيانات وتحليلها في وقت حدوثها لإنتاج إحصاءات لمعلقي التلفاز والصحفيين وكذلك لفريق التحرير الخاص بالمنصة الرقمية.

تقول Alexandra Willis: "منحتنا شركة IBM هذا العام ميزة لم نكن نملكها من قبل، وذلك باستخدام تقنية تدفق البيانات لتزويد فريق التحرير لدينا بنظرة فورية للمعالم البارزة والأخبار العاجلة".

وتضيف: "وقد راقب النظام تلقائيًا تيارات البيانات القادمة من جميع الملاعب التسعة عشر، وكلما حدث شيء مهم مثل تحقيق Sam Groth لثاني أسرع خدمة في البطولات، يتم إخبارنا بذلك على الفور. وخلال ثوانٍ، تمكنا من جلب هذه الأخبار إلى جمهورنا الرقمي ومشاركتها على الشبكات الاجتماعية لزيادة عدد الزيارات إلى موقعنا.

"وقد كانت القدرة على التقاط اللحظات المهمة والكشف عن الروايات الصريحة في البيانات، بشكل أسرع من أي شخص آخر، أمرًا أساسيًا. فإذا كنت ترغب في تجربة انفعالات بطولة The livehips، فإن أفضل ما يمكن أن يكون موجودًا في شخص ما هو متابعة الحدث على wimbledon.com".

استغلال قوة اللغة الطبيعية:

وقد خضعت للتجربة هذا العام قدرة أخرى، وهي استخدام تقنيات معالجة اللغات الطبيعية NLP الخاصة بشركة IBM وذلك للمساعدة في تطوير المكتبة الضخمة لتاريخ التنس في AELTC للحصول على معلومات ذات سياق جيد ومثيرة للاهتمام. وقد قام الفريق بتدريب IBM Watson ™ Engagement Advisor على استيعاب هذه المجموعة الغنية من البيانات غير الهيكلية واستخدامها للرد على الاستفسارات من المكتب الصحفي.

كما تم ربط نفس واجهة معالجة اللغات الطبيعية (NLP) أيضًا بقاعدة بيانات منظمة شاملة لإحصائيات المطابقة، والتي يعود تاريخها إلى أول بطولة والتي كانت عام ١٨٧٧، والذي يجيب بشكل فوري على كل من الأسئلة الأساسية والاستفسارات الأكثر تعقيدًا.

يقول Mick Desmond: «وقد أظهرت تجربة Watson قدرًا كبيرًا من الإمكانات. وفي العام القادم، وكجزء من عملية التخطيط السنوي للابتكار فإننا سننظر في كيفية استخدامنا لها على نطاق أوسع - حيث نسعى في النهاية إلى منح المعجبين فرصة أكبر للوصول إلى هذا المصدر الغني بمعلومات التنس بشكل مذهل».

الارتقاء للسحاب:

لقد تمت استضافة البيئة الرقمية بالكامل من قِبل IBM في Cloud Hybrid الخاص بها. حيث استخدمت IBM تقنيات النمذجة المتطورة للتنبؤ بالذروات في الطلب على أساس الجدول الزمني، وشعبية كل لاعب، والوقت من اليوم، والعديد من العوامل الأخرى -مما مكنها من تخصيص موارد السحابة بشكل ديناميكي لكل جزء من المحتوى الرقمي وضمان تجربة سلسة لملايين الزائرين حول العالم.

وبالإضافة إلى منصة السحابة الخاصة القوية التي دعمت البطولة لعدة سنوات، استخدمت IBM أيضًا سحابة SoftLayer® المنفصلة لاستضافة مركز قيادة ويمبلدون الاجتماعي وأيضًا توفير سعة إضافية لتكميل بيئة السحابة الرئيسة خلال أوقات الذروة للطلب.

وتعتبر مرونة بيئة السحابة أمرًا أساسيًا؛ نظرًا لأن المنصات الرقمية للبطولة تحتاج إلى أن تكون قادرة على التوسع بكفاءة من خلال عامل يزيد عن ١٠٠ في غضون أيام مع تزايد الفائدة قبل المباراة الأولى في الملعب الرئيس.

المحافظة على ويمبلدون آمنة ومُحْكَمَة:

يُعد الأمان على الإنترنت هو مصدر القلق الرئيس في الوقت الحاضر لجميع المنظمات. وبالنسبة للفعاليات الرياضية الكبرى على وجه الخصوص، فإن سمعة العلامة التجارية هي كل شيء - وبينما يراقب العالم، فإنه من المهم بشكل خاص تجنب

الوقوع كضحية سهلة للجريمة الإلكترونية cyber-crime. ولهذه الأسباب، يلعب الأمان دورًا حيويًا في شراكة IBM مع AELTC.

والجدير بالذكر أنه خلال الأشـهر الخمسـة الأولى من عام ٢٠١٥، اكتشـفت أنظمة أمان IBM زيادة بنسـبة ٩٤٪ في الأحداث الأمنية على البنيـة التحتية wimbledon.com، مقارنةً بالفـترة نفسـها من عـام ٢٠١٤. ونظرًا للتهديـدات الأمنية - وخاصة هجـمات الحرمان من الخدمة الموزعة (DDoS) - أصبحت IBM أكثر انتشارًا بشكل مستمر، حيث تعمل على زيادة تركيزها من أجل توفير مستويات رائدة في صناعة الأمن للمنصة الرقمية الكاملة لـ AELTC.

وقد مكنت مجموعة كاملة من منتجات IBM الأمنية، بما في ذلك IBM QRadar® SIEM وIBM Preventia Intrusion Prevention، بطولة هذا العام من العمل بسلاسة وأمان، ومنصة رقمية لتقديم تجربة مستخدم عالية الجودة في جميع الأوقات.

أسر القلوب والعقول:

كان نجـاح النظام الرقمـي الجديد لعام ٢٠١٥ - بدعم من تقنيات سـحابة IBM وتقنيات التحليلات وتقنيات الجوال وبالإضافة إلى التقنيات الاجتماعية والأمنية - فوريًا وكامـلًا. ولم يتـم تحقيق أهداف إجـمالي الزيارات والزوار الفريدين فحسـب، بل تم تجاوزها أيضًا. إن تحقيق ٧١ مليون زيارة و٥٤٢ مليون مشاهدة من ٢١,١ مليون جهاز فريـد من نوعه يُظهر نجـاح المنصة في جذب جمهور أكبر مـن أي وقت مضى، وإبقاء هؤلاء المشاهدين مشاركين طوال البطولة.

تقول Alexandra Willis: "بشـكل عام حققنا زيادة في الزيارات بنسبة ١٣٪ من أجهزة تجاوزت الأجهزة في عام ٢٠١٤ بنسـبة ٢٣٪، مقارنـةً بعام ٢٠١٤، وكان النمو في اسـتخدام wimbledon.com على الجوّال أكثر إثارة للإعجاب". وتضيف قائلة: "لقد شـهدنا نموًا بنسـبة ١٢٥٪ في الأجهزة الفريدة على الجوّال ونموًا بنسـبة ٩٨٪ في إجمالي الزيارات ونمو بنسبة ٧٩٪ في إجمالي عدد مشاهدات الصفحة".

ويختتم Mick Desmond قائلًا: «تُظهر النتائج أنه في عام ٢٠١٥، فزنا في الاستحواذ على قلوب وعقول المشـجعين. فقد يكون لدى الناس الصحف المفضلة ومواقع الويب الرياضية التي يزورونها لمدة ٥٠ أسـبوعًا من السـنة - غير أنهم بدلًا من ذلك قد جاءوا إلينا لمدة أسـبوعين، وهذا دليل على الجودة المطلقـة للتجربة التي يمكن أن نقدمها - واسـتغلال ميزاتنا الفريدة لجعلهم أقرب إلى العمل من أي قنوات تواصل أخرى. وقد سـاعدت القدرة على التقاط وتوصيل المحتويات ذات الصلة فور وقوعها معجبينا على تحقيق مزيد من الاستمتاع بالبطولات أكثر من أي وقت مضى".

أسئلة للمناقشة:

١- كيف استخدمت ويمبلدون قدرات التحليلات لتعزيز تجربة المشاهدين؟

٢- ما هي التحديات والحل المقترح والنتائج التي حصلوا عليها؟

Source: IBM Case Study. Creating a unique digital experience to capture the moments that matter. http://www-03.ibm.com/software/businesscasestudies/us/en/corp?synkey=D140192K15783Q68 (accessed May 2016).

تطبيقات تحليل المشاعر:

بالمقارنة مع الأســاليب التقليدية لتحليل المشــاعر، والتي كانت تستند إلى مسح أو مجموعة تركيز مركزة ومكلفة وتستهلك الكثير من الوقت (وبالتالي مدفوعة من عينة صغيرة من المشاركين)، فإن الوجه الجديد لتحليل المشــاعر القائم على تحليلات النصوص هو كاسر للحد. تعمل الحلول الحالية على جعل عمليات جمع البيانات وترشيحها وتصنيفها وتجميعها على نطاق واسع عملية آلية، وذلك عبر تقنية البرمجة معالِجة اللغات الطبيعية NLP وتقنيات التنقيب في البيانات التي تتعامل مع كل من المعلومات الواقعية والذاتية. وقد يكون تحليل المشاعر التطبيق الأكثر شيوعًا لتحليــلات النصوص، والاســتفادة من مصادر البيانات مثل التغريدات، ومشــاركات Facebook، والمجتمعات على الإنترنت، ولوحات المناقشــة، وســجلات الويب، ومراجعات المنتجات، وسجلات مركز الاتصال والتســجيلات، ومواقع تقييم المنتجات، وغرف الدردشــة، ومنافذ مقارنة الأســعار وســجلات محــرك البحث ومجموعات الأخبار. وتهــدف التطبيقات التالية لتحليل المشــاعر إلى توضيح القوة والتغطية الواسعة لهذه التقنية.

صوت العميل (VOC): يعتبر صوت العميل (VOC) جزءًا لا يتجزأ من أنظمة إدارة علاقات العمــلاء التحليليــة CRM وأنظمة إدارة العملاء. وباعتبار تحليل المشــاعر أداة تمكين لـ VOC، فإنه يمكن لتحليل المشاعر الوصول إلى مراجعات منتجات وخدمات الشركة (سواء بشكل مستمر أو بشــكل دوري) لفهم وإدارة شــكاوى العملاء ومدحهم بشــكل أفضل. فعلى سبيل المثال، قد تكتشف شركة إعلانات / تسويق، صورًا متحركة أو مشاعر سلبية تجاه فيلم على وشك الافتتاح في المســارح (استنادًا إلى مقطوراتها trailers)، وتقوم بتغيير تركيبة المقطورات trailers وإستراتيجية الإعلان بسرعة (على جميع وســائل التواصل) للتخفيف من تأثير الســلبية. وبالمثل، يمكن لشركة برمجيات أن تكتشــف الطنين الســلبي فيما يتعلق بالثغرات الموجودة في منتجها الذي تم إطلاقه حديثًا في وقت مبكر بما فيه الكفاية؛ لإصدار رقع وتصحيحات سريعة للتخفيف من حدة الموقف.

وفي كثـير من الأحيان، يكون تركيز صـوت العميل (VOC) على العملاء الأفراد، واحتياجاتهم المتعلقة بالخدمات والدعم ورغباتهم ومشكلاتهم. ويقوم VOC بتجميع البيانات من المجموعة الكاملة لنقاط اتصال العملاء، بما في ذلك رسـائل البريد الإلكتروني والاستطلاعات وتسجيلات مركز الاتصال / التسجيلات، وترحيل وسائل التواصل الاجتماعي، ومطابقة أصوات العملاء مع المعاملات (الاستفسارات والمشـتريات والعوائد)، والملفات الشـخصية للعملاء الفردية التي تم التقاطها في مؤسسـة أنظمة التشغيل. ويُعد VOC، والتي يقوم تحليل المشـاعر بقيادتها، عنصرًا أساسيًّا في مبادرات إدارة تجربة العملاء، والذي يمثل الهدف في إنشاء علاقة ودية مع العميل.

صوت السـوق (VOM): يدور VOM حول فهـم الآراء والاتجاهات المجمعة. فهي تتعلق بمعرفة ماهية أصحاب المصلحة - العملاء، العملاء المحتملين، المؤثرين، أيًّا كان - بشـأن منتجاتك وخدماتك (ومنافسيك). ويساعد تحليل VOM - الذي يتم إجراؤه جيدًا - الشركات على تطوير الذكاء التنافسي وتطوير المنتجات ووضعها.

صوت الموظف (VOE): كان VOE تقليديًّا يقتصر على دراسـة رضا الموظفين. وتُعد تحليلات النص بشكل عام (وتحليل المشاعر على وجه الخصوص) تمكينًا كبيرًا لتقييم VOE. كما يعد استخدام البيانـات النصية الغنية المعتمدة على الرأي طريقة فعالة وكفء للاسـتماع إلى ما يقوله الموظفون. وكما نعلم جميعًا، فإن الموظفين السعداء يعملون على تمكين جهود تجربة العملاء وتحسين رضاهم.

إدارة العلامـات التجاريـة: تركز إدارة العلامة التجارية على الاسـتماع إلى وسـائل التواصل الاجتماعي، حيث يمكن لأي شـخص (العملاء السـابقين / الحاليـين / المحتملين، وخبراء الصناعة، والسلطات الأخرى) نشر الآراء التي يمكن أن تضر بسمعتك أو تعززها. ويقدم عدد من الشركات المبتدئة التي تم إطلاقها حديثًا خدمات إدارة العلامات التجارية القائمة على التحليلات للآخرين. وتُعـد إدارة العلامة التجارية هي المنتج، والشركـة (بدلًا من العملاء) هي التي تركز عليها. فهي تحاول تشكيل المفاهيم بدلًا من إدارة الخبرات باستخدام تقنيات تحليل المشاعر.

الأسـواق المالية: كان التنبؤ بالقيم المستقبلية للأفراد (أو مجموعة من الأسهم) مشكلة مثيرة للاهتمام ويبدو أنها غير قابلة للحل. وما يجعل السـهم (أو مجموعة من الأسهم) يتحرك صعودًا أو هبوطًا هو أي شيء غير أنه يتطلب إدراك دقيق. ويعتقد الكثيرون أن سـوق الأسـهم مدفوعة في معظمها بالمشـاعر، مما يجعلها غير عقلانية (خاصة بالنسـبة لحركات الأسـهم قصيرة الأجل). ولذلك؛ فقد اكتسب استخدام تحليل المشاعر في الأسواق المالية شعبية كبيرة. ويبدو التحليل الآلي لميل السوق باستخدام وسائل التواصل الاجتماعي والأخبار والمدونات ومجموعات النقاش طريقة مناسبة لحساب تحركات السوق. فإذا تم ذلك بشكل صحيح، فإنه يمكن أن يحدد تحليل المشاعر تحركات الأسهم قصيرة الأجل بناءً على الضجة في السوق، مما قد يؤثر على السيولة والتداول.

السياسة: كما نعلم جميعًا، فإن الآراء مهمة للغاية في السياسة. ونظرًا لأن النقاشات السياسية تهيمن عليها الاقتباسات والسخرية والإشارات المعقدة للأشخاص والمنظمات والأفكار، فإن السياسة هي واحدة من أصعب المجالات التي يمكن أن تكون مثمرة لتحليل المشـاعر. ومن خلال تحليل المشاعر في المنتديات الانتخابية، يمكن للمرء أن يتوقع من هو أكثر احتمالًا للفوز أو الخسارة. كما يمكن أن يساعد تحليل المشاعر في فهم ما يفكر فيه الناخبون ويمكنه توضيح موقف المرشح بشأن القضايا. وبالإضافة إلى ما سبق، فإن تحليل المشاعر يمكنه أن يساعد المنظمات السياسية والحملات ومحللي الأخبار على فهم أفضل للقضايا والمواقف الأكثر أهمية بالنسـبة للناخبين. وقد تم تطبيق التقنية بنجاح من قِبل كلا الطرفين في الحملات الانتخابية الرئاسية الأمريكية لعامي ٢٠٠٨ و٢٠١٢.

المخابـرات الحكوميـة: المخابـرات الحكومية هي تطبيق آخر تم اسـتخدامه من قِبل وكالات المخابرات. فعلى سـبيل المثال، تم اقتراح إمكانية مراقبـة مصادر الزيادة في الاتصالات العدائية أو السـلبية. ويمكن لتحليل المشـاعر أن يسـمح بالتحليل التلقائي للآراء التي يقدمها الأشخاص حول مقترحات السياسـات أو التنظيم الحكومي المعلقة. وعلاوةً على ذلك، فقد تكون مراقبة الاتصالات لاكتشاف ارتفاع المشاعر السلبية مفيدة لوكالات مثل وزارة الأمن الداخلي.

مجالات أخرى مثيرة للاهتمام: يمكن استخدام مشاعر العملاء لتصميم مواقع التجارة الإلكترونية بشـكل أفضل (اقتراحات المنتج، الإعلان عن البيع / البيع المتقاطع)، ووضع الإعلانات في مكان أفضل (مثـل: وضع إعلانات ديناميكية للمنتجات والخدمات التي تنظر في المشـاعر في الصفحة التي يقوم المسـتخدم بتصفحهـا)، وإدارة الرأي أو مراجعة محـركات البحث الموجـه (أي موقع ويب لتجميع الآراء، كبديـل لمواقع مثل Epinions، وتلخيص ملاحظات المسـتخدمين). ويمكن أن يسـاعد تحليل المشاعر على تصفية البريد الإلكتروني من خلال تصنيف وترتيب الرسائل حسب أولوياتها (فمثلًا، يمكنه اكتشـاف رسائل البريد الإلكتروني السلبية وإرسـالها إلى مجلد مناسب)، بالإضافة إلى تحليل الاقتباس، حيث يمكنه تحديد ما إذا كان المؤلف يستشهد بجزء من العمل كدليل دامغ أو بحث مرفوض.

عملية تحليل المشاعر:

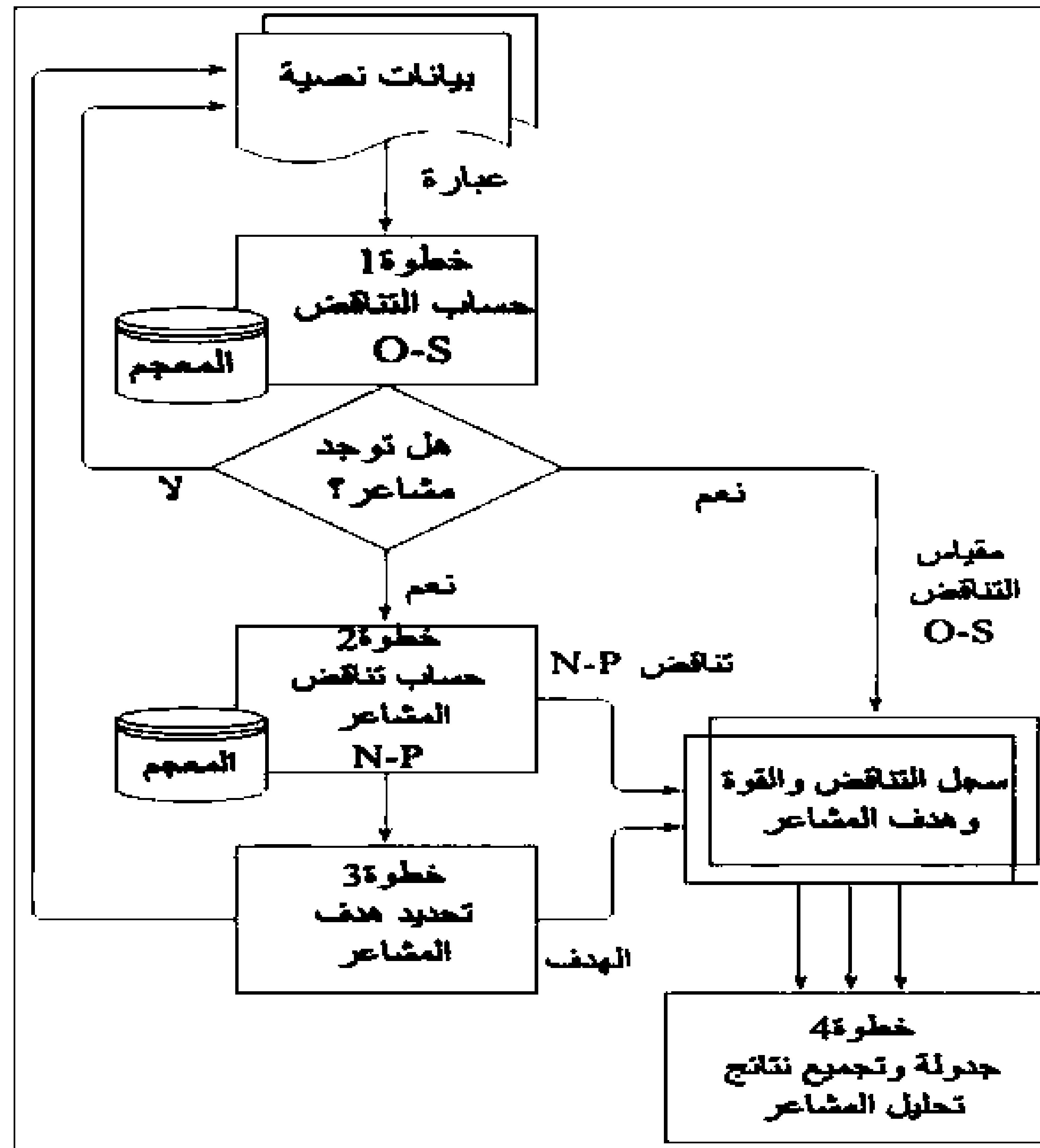
لا توجد عملية قياسـية متاحة بسـهولة لإجراء تحليل المشـاعر، وذلك بسـبب تعقيد المشـكلة (المفاهيم الأساسـية، التعبيرات في النص السـياق الذي يتم فيه التعبير عـن النص، إلخ). ومع ذلك، وبناءً على العمل المنشور في مجال تحليل الحساسية حتى الآن (على كلٍ من طرق البحث ومجموعة التطبيقات)، فإن العملية تبدو متعددة النقاط وبسيطة ومنطقية. وكما هو موضح في الشكل (٥-٩)، والذي هو عبارة عن منهجية مناسبة لتحليل المشاعر، فإن هذه الخطوات المنطقية تكرارية (أي ردود الفعل، التصحيحات، التكرارات هي جزء من عملية الاكتشاف) وتجريبية في الطبيعة، وبمجرد الانتهاء

من الجمع والتركيب، فإنها تكون قادرة على إنتاج البصيرة المرغوبة حول الآراء في مجموعة النص.

- الخطوة ١- الكشف عن المشاعر: بعد استرجاع الوثائق النصية وإعدادها، فإن المهمة الرئيسة الأولى في تحليل الحساسية هي الكشف عن الموضوعية. والهدف هنا هو التفريق بين الحقيقة والرأي، والتي يمكن اعتبارها تصنيف نصٍ موضوعي أو ذاتي. كما يمكن تمييزها بحساب O-S (الموضوعية Objectivity - تناقض الموضوعية Subjectivity Polarity، والتي يمكن تمثيلها بقيمة رقمية تتراوح من ٠ إلى ١). فإذا كانت قيمة الموضوعية قريبة من ١، فلا يوجد رأي لي (أي أنها حقيقة)؛ ولذلك ترجع العملية وتستخرج البيانات التالية للنص لتحليلها. عادةً ما يستند اكتشاف الرأي على دراسة الصفات في النص. فعلى سبيل المثال، يمكن تحديد تناقض "ما يعد عملًا رائعًا" بسهولة نسبية بالنظر إلى الصفة.
- الخطوة ٢- تصنيف التناقض N-P: المهمة الرئيسة الثانية هي تصنيف التناقض. وبالنظر إلى نص معين برأي مقنع، فإن الهدف هو تصنيف الرأي على أنه ينـدرج تحت إحدى عاطفتين متعارضتين، أو تحديد موقعه على الدوام بين هاتين المتناقضتين (Pang & Lee, 2008). فعند النظر إليه على أنه عنصر ثنائي، فإن تصنيف الأقطاب هو مهمة التصنيف الثنائي لوضع علامة على وثيقة الرأي كإعراب إما عن رأي إيجابي شامل أو سلبي شامل (مثل: الإعجاب أو الإبهام). وبالإضافة إلى تحديد تناقض N-P، فإنه يجب أيضًا أن يكون المرء مهتمًا بتحديد قوة المشاعر (على عكس الإيجابي فقط، قد يتم التعبير عنه على أنه أقل من المعتاد mildly، أو معتدل moderately أو قوي، أو إيجابي للغاية). وقد تم إجراء معظم هذا البحث حول مراجعات المنتجات أو الأفلام حيث تكون تعريفات "الإيجابية" و"السلبية" واضحة تمامًا. وهناك بعض المهام الأخرى، مثل تصنيف الأخبار على أنها "جيدة" أو "سيئة". فعلى سبيل المثال، قد تحتوي إحدى المقالات على أخبار سلبية بدون استخدام أي كلمات أو عبارات ذاتية بشكل صريح. علاوةً على ذلك، تظهر هذه الفئات عادةً مختلطة عندما تعبر الوثيقة عن المشاعر الإيجابية والسلبية. وعندئذٍ، فإن هذه المهمة يمكنها تحديد المشاعر الرئيسة (أو المسيطرة) للوثيقة. ومع ذلك، فإنه بالنسبة للنصوص المطولة، قد يلزم القيام بمهام التصنيف على عدة مستويات: المصطلح، العبارة، الجملة، وربما مستوى الوثيقة. وبالنسبة لهؤلاء، فإنه من الشائع استخدام مخرجات مستوى واحد كمدخلات للطبقة الأعلى التالية. وفي القسم التالي سوف نتناول شرح العديد من الطرق المستخدمة لتحديد التناقض وكذلك أوجه القوة في التناقض.
- الخطوة ٣- تحديد الهدف: الهدف من هذه الخطوة هو تحديد هدف المشاعر المُعبَّر عنها بدقة (على سبيل المثال: شخص، منتَج، حدث). وتعتمد صعوبة هذه المهمة بشكل كبير على مجال التحليل. وعلى الرغم من أنه من السهل عادةً تحديد الهدف بدقة لمراجعات المنتج أو الفيلم

نظرًا لأن المراجعة مرتبطة مباشرة بالهدف، فقد يكون الأمر صعبًا جدًا في النطاقات الأخرى. فعلى سبيل المثال، لا يحتوي النص طويل الأمد، مثل صفحات الويب والمقالات الإخبارية والمدوّنات دائمًا على موضوع محدد مسبقًا تم تعيينه، وكثيرًا ما يذكر العديد من الموضوعات، والتي يمكن استنتاج أي منها على أنه الهدف. وفي بعض الأحيان يكون هناك أكثر من هدف واحد في جملة المشاعر، وهذا هو الحال في النصوص المقارنة؛ حيث تقوم جملة نسبية ذاتية بطلب الأشياء حسب ترتيب التفضيل - مثل: "هذا الحاسب المحمول أفضل من الحاسب المكتبي الخاص بي". ويمكن تحديد هذه الجمل باستخدام الصفات والحالات المقارنة (أكثر، أقل، أفضل، أطول)، والصفات الفائقة (معظم، والأقل والأفضل)، وغيرها من الكلمات (مثل: نفس الشيء، اختلف، فاز، فَضَّل). وبمجرد استرداد الجمل، يمكن وضع الأشياء بترتيب أكثر تمثيلًا لمزاياها، كما هو موضح في النص.

- الخطوة ٤- الجمع والتجميع: بمجرد تحديد ومعالجة مشاعر جميع نقاط البيانات النصية في الوثيقة، يتم تجميعها وتحويلها في هذه الخطوة إلى قياس واحد لكل الوثيقة. وقد يكون هذا التجميع بسيطًا مثل تلخيص نقاط الضعف ونقاط القوة لكل النصوص، أو معقدًا مثل استخدام تقنيات التجميع الدلالي من معالجة اللغات الطبيعية NLP للوصول إلى المشاعر النهائية.



شكل ٥-٩: عملية متعددة الخطوات لتحليل المشاعر

طرق تحديد التناقض:

كما ذكرنا في القســم الســابق، يمكن تحديد التناقض على مستوى الكلمة أو العبارة أو الجملة أو الوثيقة. ويُعد المســتوى الأكثر دقة لتحديد التناقض هو على مســتوى الكلمة. وبمجرد أن يتم تحديد التناقض على مســتوى الكلمة، فإنه يمكن ضمها إلى المســتوى الأعلى التالي، ثم إلى المستوى التالي حتى يتم الوصول إلى مستوى التجميع المطلوب من تحليل المشاعر. ويبدو أن هناك تقنيتين سائدتين تُستخدمان لتحديد التناقض على مستوى الكلمة / المصطلح، ولكل منهما مميزاته وعيوبه:

١- اســتخدام المعجــم كمكتبة مرجعية (ويتم تطويــره إما يدويًا أو تلقائيًا، بواســطة فرد لمهمة محددة أو قد يتم تطويره بواسطة مؤسسة للاستخدام العام).

٢- اســتخدام مجموعة من وثائق التدريب كمصــدر للمعرفة حول تناقض المصطلحات في نطاق معين (أي استحداث نماذج تنبؤية من الوثائق النصية المعترف بها).

استخدام المعجم:

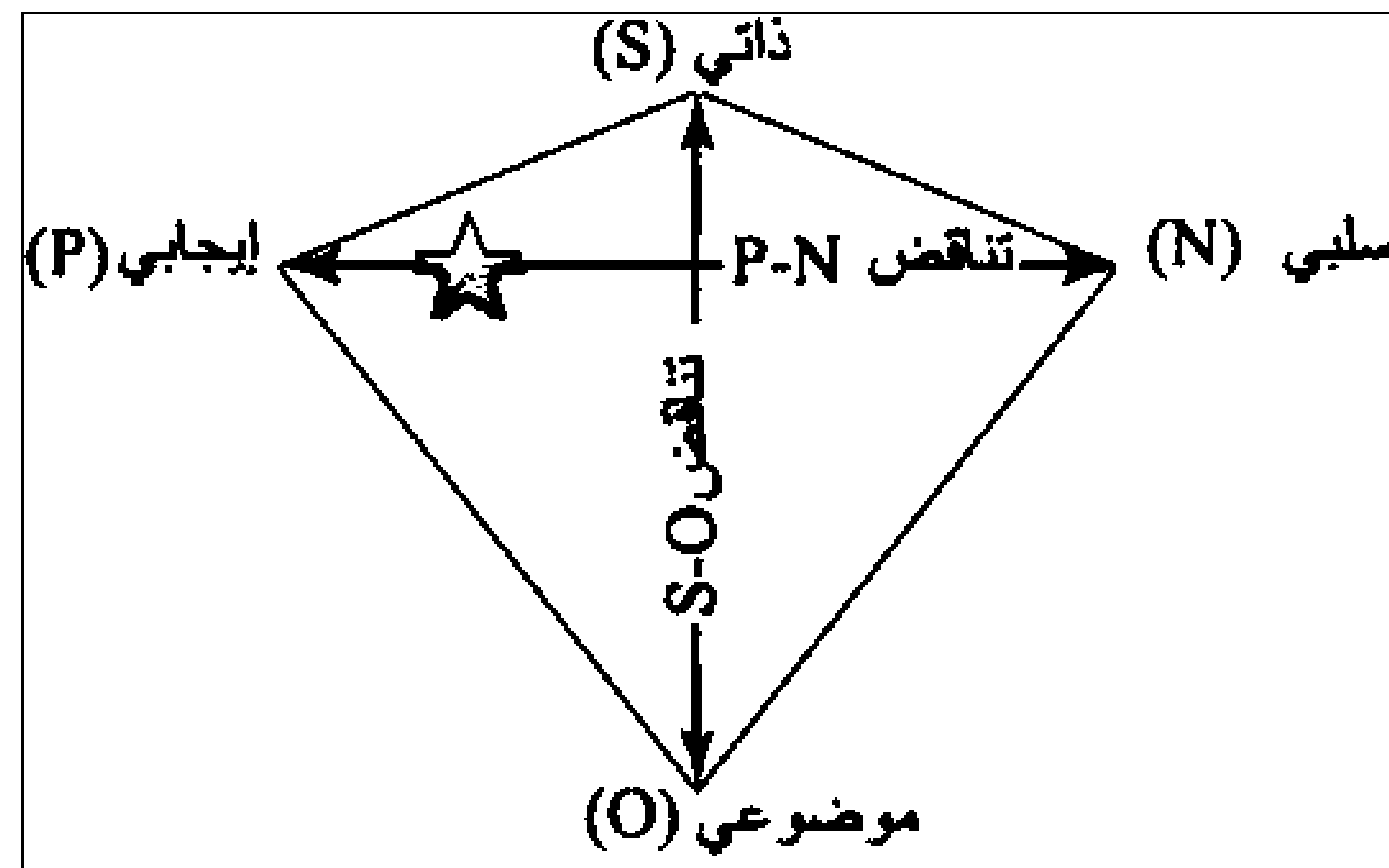
المعجم هو في الأســاس كتالوج لكلمات لغة معينــة ومرادفاتها ومعانيها. وهناك العديد من المعاجم ذات الأغراض العامة التي تم إنشاؤها للغة الإنجليزية، بالإضافة إلى وجود معاجم للعديد من اللغات الأخرى. وتســتخدم معاجم الأغراض العامة في الغالب لإنشــاء مجموعة متنوعة من المعاجم ذات الأغراض الخاصة لاســتخدامها في مشــاريع تحليل المشــاعر. ولعــل المعجم الأكثر انتشــارًا للأغــراض العامة هو WordNet، والذي تم إنشــاؤه في جامعــة Princeton، والذي تم توســعته واســتخدامه من قِبل العديد من الباحثين والممارسين لأغراض تحليل المشاعر. وكما هو موضح في موقع WordNet على الويب (wordnet.princeton.edu)، فهو قاعدة معجمية كبيرة للغة الإنجليزية، بما في ذلك الأســماء والأفعال والصفات والظروف، والتي تم ضمها في مجموعات من المرادفــات المقربة (ويطلق عليها synsets)، والتي يعبر كل منها عن مفهوم متميز. وتترابط synsets عن طريق وسائل مفاهيمية دلالية وعلاقات معجمية.

وقد تم إنشــاء ملحق مثير للاهتمام لمعجم WordNet بواسطة Esuli وSebastiani (2006) حيــث قاما بإضافــة تناقض (إيجابي - ســلبي؛ P-N) وموضوعيــة (ذاتي - موضوعي؛ S-O) لكل مصطلح في المعجم. ولتسمية كل مصطلح، يقومان بتصنيف المجموعة من المترادفات التي ينتمي إليها هذا المصطلح باســتخدام مجموعة ثلاثية التصنيف (وهــي مقياس يربط كل موضوع تمامًا واحدًا من ثلاثة عناوين)، وكل واحد منهم قادر على تقرير ما إذا كانت synset إيجابية أم سلبية

أم موضوعيــة. وتتراوح الدرجات الناتجة مــن ٠,٠ إلى ١,٠، مما يعطي تقييمًا متدرجًا لخصائص المصطلحات المتعلقة بالرأي. ويمكن تلخيصها بصورة مرئية كما في الشكل (٥-١٠). وتمثل أضلاع المثلث التصنيفات الثلاثة (الإيجابية والسلبية والموضوعية). ويمكن أن يتم تحديد موقع المصطلح في هذه المساحة كنقطة، تمثل المدى الذي ينتمي إليه كل واحد من هذه التصنيفات.

ويتم استخدام منهجية توسيع مماثلة لإنشاء SentiWordNet، وهو معـجـم متاح للجمهور تــم تطويره خصيصًا لأغراض التنقيب في الرأي (تحليل المشــاعر). ويُعيّن SentiWordNet لكل مجموعة من الدرجات الثلاث من معنويات WordNet: الإيجابية والسلبية والموضوعية. ويمكن الاطلاع على المزيد حول SentiWordNet في sentiwordnet.isti.cnr.it.

وهنــاك ملحق آخر لـ WordNet هو WordNet-Affect، وقد تم تطويره بواســطة كل من Strapparava وValitutti (2004)، حيث قاما بتصنيف معادلات WordNet باستخدام تسميات فعالة تمثل فئات عاطفية مختلفة مثل العاطفة والحالة الإدراكية والموقف والشعور. كما استخدما WordNet أيضًا بشــكل مباشر في تحليل المشــاعر. فمثلًا، قام كل من Kim وHovy (2004)، و Liu وHu وCheng (2005) بتطويــر معاجم للمصطلحات الإيجابية والســلبية من خلال البدء بقائمة صغيرة من مصطلحات "البذور" للمترادفات المتعارف عليها (مثل: nice, like, love) وبعد ذلك استخدما خاصيتَي تضاد وترادف المصطلحات لتجميعها في أي فئة من فئات التناقض.



شكل ٥-١٠: تمثيل بياني لعلاقتَي تناقض كل من P-N وS-O

استخدام مجموعة من وثائق التدريب:

من الممكن إجراء تصنيف للمشــاعر باســتخدام التحليل الإحصائي وأدوات التعليم الآلي التي تتمتــع بميزة كمية الموارد الضخمة من الوثائق المتاحة (يدويًا عن طريق الحواشي أو باســتخدام

نظــام نجمة / نقطة). وقد تم اســتخدام مواقع الويب مثــل Amazon، وC-NET، وeBay، و RottenTomatoes، وInternet Movie Database على نطاق واسع كمصادر للبيانات المشروحة، من أجل استعراض المنتجات. ويوفر نظام النجمة (أو الطماطم سابقًا) علامة واضحة عن التناقض الكامل في الاستعراض ككل، وغالبًا ما يتم استخدامه كمعيار ذهبي لتقييم الخوارزمية.

وتتوفر مجموعة متنوعة من البيانات النصية المصنفة يدويًا من خلال جهود التقييم مثل مؤتمر RETRIETER، وNII Test Collection لأنظمــة IR، وعــبر منتدى تقييم اللغات. وتنتج هذه المجموعة من البيانات جهودًا غالبًا ما تكون بمثابة معيار في مجتمع التنقيب في النص، بما في ذلك لباحثي تحليل المشاعر. كما أنتج الباحثون الفرديون ومجموعات البحث العديد من مجموعات البيانات المثــيرة. وتعرض الفكرة التطبيقية (٥-٢) قائمة ببعضها. وبمجرد الحصول على مجموعة بيانات نصية مصنفة بالفعل، فإنه يمكن استخدام مجموعة متنوعة من النمذجة التنبؤية وغيرها من خوارزميات تعلم الآلة لتدريب مصنفي المعنويات. وتتضمن بعض الخوارزميات الأكثر شيوعًا المستخدمة لهذه المهمة الشــبكات العصبية الاصطناعية، وآلة المتجهات الداعمة، والجار الأقرب، والخلايا البسيطة، وأشجار القرار، وتعظيم التقديرات على أساس المجموعة.

رؤية فنية ٥-٢

مجموعات بيانات نصية كبيرة للتنقيب التنبؤي في النص وتحليل المشاعر

نصوص الكونغرس حول Floor-Debate: نشرها كل من Thomas، وPang وLee (2006)؛ وتحتوي على خطابات سياسية تم تصنيفها لتوضيح ما إذا كان المتحدث يدعم أو يعارض التشريع الذي تمت مناقشته.

Economining: نشرتها مدرســة Stern في جامعة نيويورك، وتتكون من تعقيبات التجار في Amazon.com.

مجموعــات بيانات عــرض الأفلام من Cornell: وقد قدمهــا كل Pang و(Lee 2008)؛ وتحتوي على ١٠٠٠ تصنيف إيجابي و١٠٠٠ تصنيف تلقائي سلبي مشتقة من عناوين مستوى الوثيقة و٥٣٣١ جملة / قُصاصة إيجابية و٥٣٣١ جملة / قصاصة سلبية.

ســتانفورد - مجموعة كبيرة لعرض الأفلام: وهي مجموعة من ٢٥٠٠٠ فيلم شديدة التناقض تتم مراجعتها للتدريب و٢٥٠٠٠ للاختبار. ويوجد بيانات إضافية للاســتخدام بــدون عنــوان. كما يتم توفير تنســيقات للنصوص الخام وحقائــب الكلمات المعالجة مسبقًا. (انظر: http://ai.stanford.edu/~amaas/data/sentiment).

مجموعة MPQA: تحتوي المجموعة ونظام استطلاع الرأي على ٥٣٥ مقالًا إخباريًا مشروحة يدويًا من مجموعة متنوعة من مصادر الأخبار التي تحتوي على عناوين للآراء والتصريحات الخاصة (المعتقدات، العواطف، المضاربات، إلخ).

عروض مطاعم متعددة الجوانب: وقام بعرضها كل من Snyder وBarzilay (2007)؛ وتحتوي على ٤٤٨٨ مراجعة مع تقييم صريح من ١ إلى ٥ لخمسة جوانب مختلفة: الطعام والجو والخدمة والقيمة بالإضافة إلى الخبرة العامة.

تحديد الاتجاه الدلالي للجمل والعبارات:

بمجـرد تحديـد الاتجاه الدلالي للكلمات الفردية، فمن المستحسـن في كثير من الأحيان تمديد ذلـك إلى العبارة أو الجملة التي تظهر بها الكلمة. وأبسـط طريقـة لتحقيق هذا التجميع هي استخدام نوع من المتوسط لاستقطاب الكلمات في العبارات أو الجمل. وعلى الرغم من أنه نادرًا ما يتم تطبيقه، إلا أن هذا التجميع يمكن أن يكون معقدًا مثل استخدام واحد أو أكثر من تقنيات تعلم الآلة لإنشاء علاقة تنبؤية بين الكلمات (وقيم تناقضها) والعبارات أو الجمل.

تحديد الاتجاه الدلالي للوثائق:

على الرغم من أن الغالبية العظمى من العمل في هذا المجال تتم في تحديد التوجه الدلالي للكلمات والعبارات / الجمل، فإن بعض المهام مثل التلخيص واسـترجاع المعلومات قد تتطلب وضع العلامات الدلالية للوثيقة بأكملها (Ramage وآخرون، ٢٠٠٩). ويتم أيضًا تحقيق التجميع إلى مستوى الوثيقة بواسطة نوع من المتوسط، على غرار الحالة في تجميع تناقض المشاعر من مستوى الكلمة إلى مستوى شـبه الجملة أو الجملة. وقد لا يكون اتجاه المشـاعر للوثيقة منطقيًـا للوثائق الكبيرة جدًا؛ ولذلك، فغالبًا ما يتم استخدامه على الوثائق الصغيرة إلى متوسطة الحجم المنشورة على الإنترنت.

أسئلة مراجعة على القسم ٥-٦:

١- ما هو تحليل المشاعر؟ وكيف يرتبط بالتنقيب في النص؟

٢- ما هي مجالات التطبيق الأكثر انتشارًا لتحليل المشاعر؟ ولماذا؟

٣- ما هي الفوائد المتوقعة والمستفادة من تحليل المشاعر في السياسة؟

٤- ما هي الخطوات الرئيسة في تنفيذ مشاريع تحليل المشاعر؟

٥- ما هما الطريقتان الشائعتان لتحديد التناقض؟ اشرح.

٥-٧ نظرة عامّة على التنقيب في الويب:

لقد غيرت شـبكة الإنترنت خطوات إجراءات الأعمال إلى الأبـد. ونظرًا لكون العالم الآن متصلًا ببعضه البعض ومترابطًا بشكل كبير وأيضًا لاتساع مجال المنافسة، تواجه الشركات اليوم بشكل متزايد فرصًـا أكبر (وهي القدرة على الوصول إلى عملاء وأسـواق لم يكـن يُعتقد من قبل إمكانية الوصول إليها) وتحديات أكبر (أي السوق التنافسي المعولم والمتغير باستمرار). فالأفراد الذين يتمتعون برؤية وقدرات للتعامل مع مثل هذه البيئة المتقلبة يسـتفيدون منها بشـكل كبير، في حين يواجه غيرهم ممن يكرهون التغيير صعوبة في البقاء في السـوق. ولم يعد الدخول على شبكة الإنترنت الآن خيارًا كماليًا، فقد أصبح أحد متطلبات العمل الآن. فالعملاء ينتظرون اسـتلام المنتجات / الخدمات من الـشركات عبر الإنترنت. ولا يتوقف الأمر على الشراء فحسـب، فهم أيضًا يشـاركون مع أصدقائهم تجاربهم ومعاملاتهم التي تتم مع الشركات.

لقد أدى النمو المتسـارع في الإنترنت وتقنياته التمكينية إلى تسـهيل عمليات إنشـاء وجمع البيانـات وتبادل البيانات / المعلومات / الآراء. كـما أن التأخيرات في الخدمة وعمليات التصنيع والشـحن والتسـليم واسـتعلامات العملاء لم تعد اليوم حوادث خاصة ويتم استقبالها والتعامل معهـا باعتبارها أمورًا ضرورية. فالشركات الناجحة هي التي تتبنّى تقنيات الإنترنت الحديثة فور ظهورهـا وتطوعها من أجل تحسـين عملياتها التجارية بحيث يمكنها التواصل بشـكل أفضل مع عملائها وفهم احتياجاتهم ورغباتهم وخدمتهم بشكل كامل وسريع. فلم يكن إطلاقًا التركيز على العملاء والحفاظ على مسـتوى رضاهم مفهومًا مهمًا للشركات كما هو في عصر الإنترنت ووسـائل التواصل الاجتماعي التي نعيشها اليوم.

تعتبر شبكة الإنترنت بمثابة مستودع ضخم للبيانات والمعلومات حول كل شيء يمكن أن يتصوره أي أحد. على سبيل المثال لا الحصر، الأعمال التجارية، الأبحاث والنظريات العلمية، مقالات في شتى المجالات، وبكميات وفيرة. وبذلك فربما تكون شـبكة الويب هي أكبر مستودع للبيانات والنصوص في العـالم، ويتزايد حجم المعلومات على الويـب بسرعة، مما يمكّن أي باحث من العثور على الكثير مـن المعلومات المثيرة لاهتمامه والوصول إلى صفحتها الرئيسـية والصفحـات الأخرى المرتبطة بها، وكذلك إلى الأشـخاص الذين لديهم روابط إلى تلك الصفحة. وبالإضافة إلى ذلك، يقوم كل زائر لأي موقـع ويب وكل بحث عـلى أي محرك بحث وكل نقرة على أي رابـط وكل معاملة على أي موقع للتجارة الإلكترونية بإنشاء بيانات إضافية. وعلى الرغم من أن البيانات النصية غير المهيكلة في شكل صفحـات الويـب والمكتوبة بلغة HTML أو XML هي المحتـوى المهيمن على الويب، فإن البنية التحتية للويب تتضمن معلومات ارتباط تشـعبي hyperlink information (روابط إلى صفحات

ويب أخرى) ومعلومات الاستخدام (سجلات تفاعلات الزوار مع مواقع الويب)، وكلها توفر بيانات غنية لاكتشاف المعرفة. كما يمكن أن يساعدنا تحليل هذه المعلومات في تحسين استخدام مواقع الويب وكذلك مساعدتنا في تعزيز العلاقات والقيمة لزوار مواقعنا الخاصة.

نخلص مما سبق إلى أن شبكة الإنترنت تتمتع بحجم هائل ومعقد، مما يجعل التنقيب في الويب ليس مهمةً سهلةً بأي وسيلة من الوسائل. كما تشكل الويب أيضًا تحديات كبيرة لاكتشاف فعال وكفء للمعرفة (Han & Kamber, 2006):

- شبكة الويب كبيرة جدًا للتنقيب الفعّال عن البيانات: إن شبكة الإنترنت كبيرة جدًا ومتنامية بسرعة بحيث يصعب حتى قياس حجمها. ولذا؛ فليس من الممكن إنشاء مستودع بيانات لتكرار وتخزين ودمج جميع البيانات على الويب، مما يجعل جمع البيانات وتكاملها أمرًا غاية في الصعوبة.
- الويب معقدٌ للغاية: يعتبر تعقيد صفحة الويب أكبر بكثير منه لصفحة في مجموعة وثائق نصية تقليدية. وتفتقر صفحات الويب إلى بنية موحدة؛ فهي تحتوي على أنماط للتأليف ومحتوى أكبر بكثير من أي مجموعة من الكتب أو المقالات أو غيرها من الوثائق النصية التقليدية.
- الويب حيويٌّ جدًا: يعتبر الويب مصدرًا للمعلومات مفعمًا بالحيوية لأبعد الحدود. ولا ينمو الويب بسرعة، بل يتم تحديث محتواه باستمرار؛ إذ يتم تحديث المدونات والقصص الإخبارية ونتائج أسواق الأسهم وتقارير الطقس والنتائج الرياضية والأسعار وإعلانات الشركات وأنواع أخرى من المعلومات بشكل منتظم على الويب.
- شبكة الويب ليست خاصة بمجالٍ ما: تخدم الويب مجموعة كبيرة من المجتمعات كما تقوم بربط المليارات من محطات العمل. يتمتع مستخدمو الويب بخلفيات واهتمامات وأغراض استخدام مختلفة تمامًا. كما قد لا يكون لدى معظم المستخدمين معرفة جيدة بهيكل شبكة المعلومات أو قد لا يكونون على دراية بالتكلفة الباهظة لعملية بحث معينة يقومون بها.
- الويب يحتوي على كل شيء: يعتبر جزء صغير فقط من المعلومات الموجودة على الويب ذا صلة أو مفيد لشخصٍ ما (أو لمهمة ما). يُقال إن ٩٩٪ من المعلومات على الويب لا فائدة منها لـ ٩٩٪ من مستخدمي الإنترنت. وعلى الرغم من أن هذا قد لا يبدو واضحًا، إلا أن شخصًا ما قد يهتم بجزء صغير فقط من الويب، بينما تحتوي بقية الويب على معلومات يراها غير مهمة بالنسبة لـه وهي قد تؤدي إلى النتائج المرجوة؛ إذ يعتبر العثور على جزء من الويب ذي صلة حقيقية بشخص ما والمهمة التي يقوم بتنفيذها مسألة شائكة فيما يخص البحث المتعلق بالويب.

دفعت كل هذه التحديات العديد من الجهود البحثية لتعزيز فعالية وكفاءة اكتشاف واستخدام أصول البيانات على شبكة الإنترنت، حيث يقوم عدد من محركات البحث على الويب

القائمة على الفهرســة بالبحث باســتمرار في الويب وتقوم بفهرسة صفحات الويب تحت كلمات رئيسية معينة. وباستخدام هذه المحركات قد يتمكن مستخدم متمرس من تحديد موقع الوثائق من خلال توفير مجموعة من الكلمات المفتاحية أو العبارات المقيدة بإحكام. ومع ذلك، فإن أي محرك بحث بسيط يعتمد فقط على الكلمات المفتاحية قد يواجه بعض المشكلات، منها:

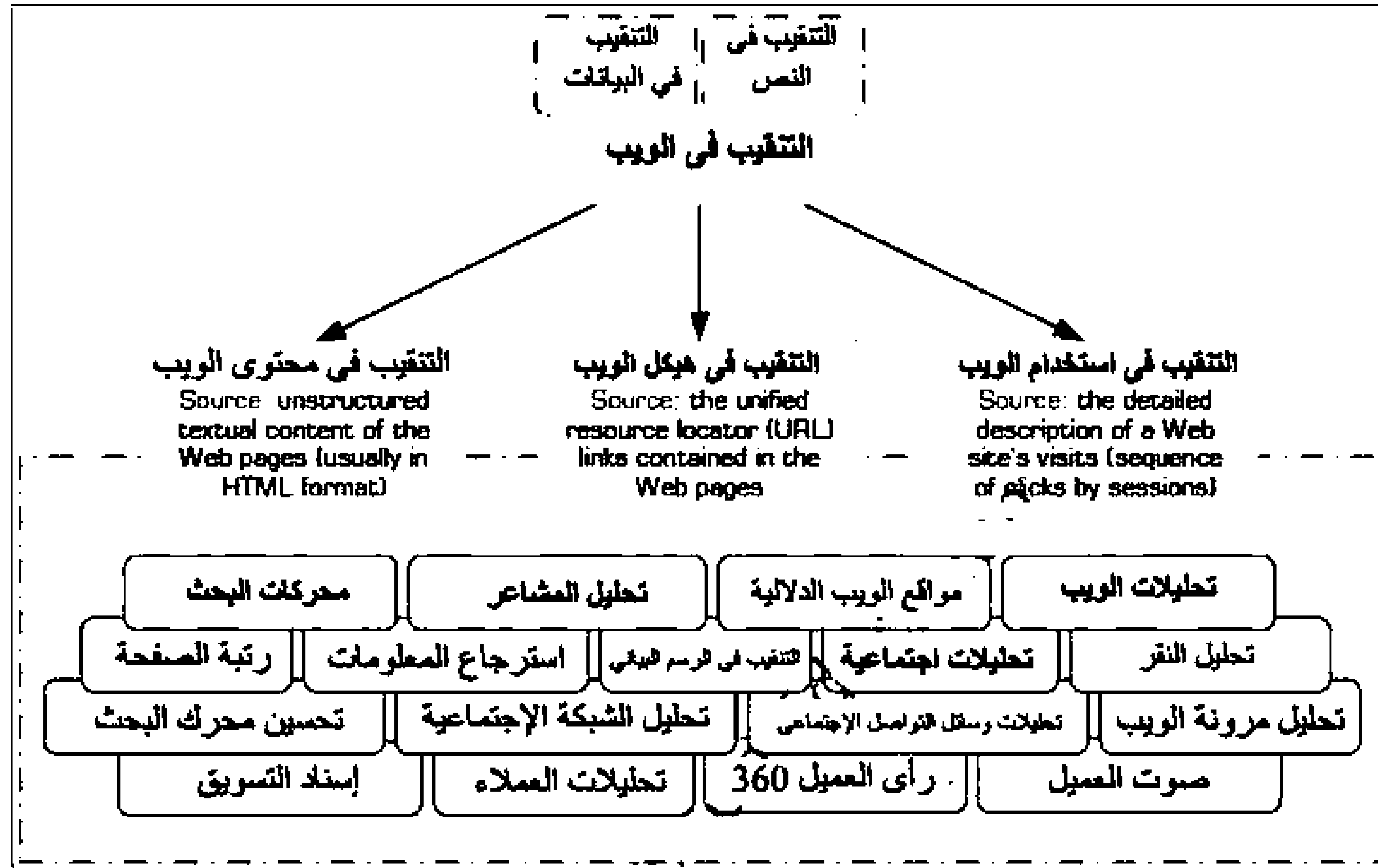
١- يمكــن لأي موضوع معروض للبحث أن يحتوي بســهولة على مئــات أو آلاف الوثائق. ويمكن أن يــؤدي ذلك إلى عدد كبير من المدخلات ذات صلة هامشــية بالموضوع والتي يتم إرجاعها بواسطة محرك البحث.

٢- قد لا تحتوي العديد من الوثائق ذات الصلة الكبيرة بالموضوع محل البحث على كلمات رئيسة دقيقة لتحددها.

كما سنقوم بتغطية أكثر تفصيلًا لاحقًا في هذا الفصل، مقارنةً بالبحث على الويب القائم على الكلــمات المفتاحية. يُعد التنقيب في الويب أحد الأســاليب البــارزة (والأكثر تحديًا) التي يمكن اســتخدامها لتعزيز قوة محركات البحث على الويب بشــكل كبــير؛ لأن التنقيب في الويب يمكنه تحديــد صفحات الويب، وتصنيف وثائق الويب، وحــل العديد من الغموض والخواطر التي يتم جمعها في محركات البحث القائمة على الكلمات المفتاحية.

يعــرف التنقيــب في الويب (أو التنقيــب في بيانات الويــب) بأنه عملية اكتشــاف العلاقات الجوهريــة (أي المعلومات المثيرة للاهتمام والمفيدة) من بيانات الويب، والتي يتم التعبير عنها في شــكل معلومات نصية أو ارتباطية أو معلومات الاســتخدام. وقد تم اســتخدام مصطلح التنقيب في الويــب Web mining لأول مرة بواســطة Etzioni (1996)؛ وتركز اليوم العديد من المؤتمرات والمجلات والكتب على التنقيب في بيانات الويب. إنه مجال متطور باستمرار من التقنية والممارسات التجارية. ويعتبر التنقيب على الويب في الأســاس نفس عملية التنقيب في البيانات التي تستخدم البيانات التي يتم توليدها عبر الويب. ويكون الهدف من ذلك هو تحويل مستودعات واسعة من المعاملات التجارية وتفاعلات العملاء وبيانات استخدام موقع الويب إلى معلومات قابلة للتطبيق (أي المعرفة) وذلك لتشــجيع اتخاذ قرارات أفضل في جميع أنحاء المؤسسة. وبسبب زيادة انتشار مصطلح التحليلات، شرع الكثيرون في الوقت الحاضر في الرجوع إلى التنقيب على شــبكة الإنترنت وتحليلات الويب. ومع ذلك، لا يعتبر هذان المصطلحان شيئًا واحدًا. وفي حين تعد تحليلات الويب في المقام الأول هي بيانات استخدام موقع الويب، فإن التنقيب على الويب يشمل جميع البيانات التي يتم توليدها عبر الإنترنت بما في ذلك بيانات المعاملات الاجتماعية وبيانات الاستخدام. وحيث

تســتهدف تحليلات الويب وصف ما حدث على موقع الويب (باستخدام منهجية تحليلية وصفية مبنية على مقاييس ومحددة مســبقًا)، يهدف التنقيب في الويب إلى اكتشــاف أنماط وعلاقات غير معروفة من قبل (باســتخدام منهجية تحليلية جديدة للتنبــؤ أو منهجية توجيهية). ومن منظور أكثر اتســاعًا، يمكن اعتبار تحليلات الويب جزءًا من التنقيب في الويب. ويعرض الشــكل (٥-١١) تصنيفًا بســيطًا للتنقيب على الويب، حيث يتم تقســيمه إلى ثلاثة مجالات رئيسة هي: التنقيب في محتــوى الويب والتنقيب في بنية الويب والتنقيب باســتخدام الويب. وفي هذا الشــكل أيضًا، يتم تحديد مصادر البيانات المستخدمة في هذه المناطق الرئيسة الثلاثة. وعلى الرغم من أن هذه المجالات الثلاثة تظهر بشــكل منفصل، كما ســنرى في القسم التالي من هذا الفصل، إلا أنها غالبًا ما تُستخدم بشكل جماعي وتكاملي لمعالجة مشكلات وفرص الأنشطة التجارية.



شكل ٥-١١: تصنيف مبسط للتنقيب في الويب

وكما يوضح الشــكل (٥-١١)، يعتمد التنقيب على الويب بشــكل كبير عــلى التنقيب في البيانات واســتخراج النصوص وأدواتها وتقنياتها التمكينيــة، والتي قمنا بتغطيتها بالتفصيل في وقت مبكر من هذا الفصل وفي الفصل الســابق (الفصل ٤). ويشــير هذا الشــكل أيضًا إلى أن هذه المناطق العامة الثلاثــة تمتــد إلى العديد من مناطــق التطبيق المعروفة جيدًا. وقد تــم شرح بعض هذه المجالات في الفصول السابقة، وسيتم تغطية بعض الجوانب الأخرى بالتفصيل في هذا الفصل.

التنقيب في محتوى الويب وهيكل الويب:

يشــير مصطلح التنقيب في محتوى الويب إلى استخراج المعلومات المفيدة من صفحات الويب؛ حيث يتم اســتخراج الوثائق بتنســيق يمكن قراءته آليًّا حتى تتمكن التقنيات الآلية من استخلاص بعض المعلومات من صفحات الويب هذه. كما يتم استخدام برامج زواحف الويب Web crawlers (ويُطلق عليها أيضًا العناكب Spiders) لقراءة محتوى موقع الويب تلقائيًّا. وقد تتضمن المعلومات التــي تم تجميعها خصائص وثائقية مشــابهة لما يتم اســتخدامه في التنقيــب في النص، ولكنها قد تتضمن أيضًا مفاهيم إضافية، مثل التسلســل الهرمي للوثائق. ويمكن استخدام عملية الجمع الآلي (أو شــبه الآلي) لمحتوى الويــب والتنقيب فيه للحصول على معلومات ذكاءية تنافســية (بما يعني جمع المعلومات المتعلقة بالمنافســين ومنتجاتهــم وخدماتهم التي يقدمونها وعملائهم). كما يمكن اســتخدامه أيضًــا لجمع المعلومات / الأخبــار / الآراء وتلخيصها وتحليل المشــاعر والجمع التلقائي للبيانــات وتنظيمها من أجــل النمذجة التنبؤية. وكمثال توضيحي لاســتخدام التنقيب في محتوى الويب كأداة لجمع البيانات بطريقة آلية، دعنا نأخذ في الاعتبار ما يلي: منذ أكثر من ١٠ ســنوات، قام كل من دكتور Sharda وDelen بتطوير نماذج للتنبؤ بالنجاح المالي لأفلام هوليوود قبل إصدارها المسرحي؛ حيث تأتي البيانات التي يستخدمونها لتدريب النماذج من عدة مواقع على الويب، لكل منها بنية صفحة هرمية مختلفة. والجدير بالذكر أن عملية جمع مجموعة كبيرة من المتغيرات على آلاف الأفلام (من عدة ســنوات ســابقة) من مواقع الويب عملية تتطلب الكثير من الوقت وكذلك عرضة للخطأ. ولذلك؛ فهي تستخدم التنقيب في محتوى الويب والعناكب Spiders باعتبارها تقنية تمكينية لجمع واختبار صحة وصلاحية وتخزين هذه القيم في قاعدة بيانات علائقية. (وفيما يخص اختبــار الصحة والصلاحية، فإذا كان عنصر البيانات المحــدد متاحًا على أكثر من موقع ويب واحد، عندئذٍ يتم التحقق من صحة القيم ضد بعضها البعض ويتم التقاط وتســجيل الانحرافات)، وبذلك فهي تضمن جودة البيانات مع توفير وقت العملية الثمين (أيام أو أسابيع).

وبالإضافــة إلى احتوائها على النصــوص، تحتوي صفحات الويب أيضًا على روابط تشــعبية تشــير صفحة واحدة إلى أخــرى hyperlinks؛ إذ تحتوي الروابط التشــعبية على قدر كبير من التعليقات التوضيحية البشرية المخفاة التي يمكن أن تساعد في الاستنتاج التلقائي لمفهوم المركزية أو السلطة. وعندما تشتمل صفحة ويب على رابط يشير إلى صفحة ويب أخرى، فيمكن اعتبار ذلــك بمثابة موافقة المطوِّر على الصفحة الأخرى. وقد يشــير التأييــد الجماعي لصفحة معينة على الويب من قِبل مطوِّرين مختلفين إلى أهمية تلك الصفحة، وقد يؤدي ذلك بشــكلٍ طبيعي إلى اكتشــاف عدة صفحــات ويب موثوقــة (Miller, 2005). لذلك؛ فإن وجــود كمية هائلة

مــن المعلومات المترابطة عــبر الإنترنت يوفر مجموعة غنية مــن المعلومات حول مدى ملاءمة محتويات الويب وجودتها وهيكلها، وتعتبر بالتالي مصدرًا غنيًّا للتنقيب في الويب.

كما يمكن اســتخدام التنقيب في محتوى الويــب لتعزيز النتائج التي تنتجها محركات البحث. وفي الواقع، قد يكون البحث هو التطبيق الأكثر انتشــارًا للتنقيب في محتوى الويب وتعديل بنية الويــب. وعادةً مــا يقودنا البحث على الويــب الذي يتم بغرض الحصول عــلى معلومات حول موضــوع معين (والذي يتم تقديمه كجملة أو مجموعة كلمات مفتاحية) إلى بضع صفحات ويب ذات صلــة بالموضوع محل البحــث، وتكون ذات جودة عالية وأيضًــا إلى عدد أكبر من صفحات الويب غير القابلة للاســتخدام. ويؤدي استخدام الفهرس المتعلق بالموضوع محل البحث استنادًا إلى الكلمات المفتاحية والصفحات الرسمية (أو بعض مقاييسها) إلى تحسين نتائج البحث وترتيب الصفحــات ذات الصلــة. وتنبع فكرة المســؤولية الفكرية (أو الصفحات الموثوقة) من اســترجاع المعلومات ســابقة النشر باســتخدام اقتباســات من مقالات المجلات لتقييم تأثير الأوراق البحثية (Miller, 2005). وبالرغــم مــن أن هذا هو أصــل الفكرة، إلا أن هنــاك اختلافات معنوية بين الاقتباسات في المقالات البحثية والروابط التشعبية hyperlinks على صفحات الويب وهي:

١- لا يمثل كل رابط تشعبي مصادقة أو موافقة (فبعض هذه الروابط يتم إنشاؤه لأغراض التصفح وبعضهــا للإعلانات المدفوعة). وعــلى الرغم من أن هذا صحيح، إذا كانت غالبية الارتباطات التشعبية من نوع المصادقة، فسيظل الرأي الجماعي سائدًا.

٢- بالنسبة للحالات التجارية والتنافسية، نادرًا ما ستشير إحدى الإدارات على صفحتها على الإنترنت إلى الصفحــات المنافســة لها في نفس المجال. فعلى ســبيل المثال، قــد لا تفضل Microsoft تضمين روابط على صفحات الويب الخاصة بها تشــير إلى المواقع الخاصة بشركة Apple؛ لأن هذا قد يعتبر بمثابة تصديق على سلطة منافسها.

٣- نادرًا ما تكون الصفحات الموثوقة وصفية بشــكل خاص. فمثلًا، قد لا تحتوي صفحة Yahoo الرئيسية على وصف ذاتي صريح بأنه في الواقع محرك بحث.

وقــد أفرز هيكل الروابط التشــعبية على الويب تصنيفًا آخر مهمًا لصفحات الويب يســمى المحور hub، وهو واحد أو أكثر من صفحات الويب التي توفر مجموعة من الروابط المشــيرة إلى صفحــات موثوقة أخرى. وقد لا تكون المحاور بارزة، وأيضًا قد يشــار إليهــا بروابط قليلة جدًا، ومع ذلك فإنها توفر روابط تشــير إلى مجموعة مــن المواقع البارزة حول الموضوع محل الاهتمام أو الدراســة. وقد يكون المحور عبارة عن قائمة بالروابط الموصى بها على الصفحات الشخصية، أو

قائمـــة بالمواقع المرجعية المقترحة على صفحة الويب الخاصة بدورة تدريبية ما، أو قائمة موارد تم تجميعها بشـــكل احترافي حول موضوع محدد. كما تلعب المحاور دورًا ضمنيًا في الحد من اتساع نطاق الســـلطات. وجوهريًا، توجد علاقة تكافلية وثيقة بين المحاور الجيدة والصفحات الموثوقة؛ فالمحاور الجيدة تكون جيدة لأنها تشير إلى العديد من السلطات الجيدة، والسلطة الجيدة تكون جيدة بســـبب الإشـــارة إليها من قِبل العديد مـــن المحاور الجيدة. وهـــذه العلاقات بين المحاور والسلطات تجعل من الممكن تلقائيًا تلقّي محتوى عالي الجودة من الويب.

إن أكثر الخوارزميات المعروفة والمشـــار إليها بشـــكل عام والمســـتخدمة في حســـاب المحاور والســـلطات هي البحث عن الموضوع الذي يُحدثه الارتباط التشعبي (HITS) والذي تم تطويره في الأصـــل مـــن قبل (Kleinberg, 1999). ومنذ ذلك الحين تم تحســـينه أكثر من مرة بواســـطة العديد من الباحثين. إن HITS عبارة عن خوارزمية تحليل رابط تقوم بتصنيف صفحات الويب باستخدام معلومات الرابط التشعبي الموجودة بداخلها. وفي سياق بحث الويب، تقوم خوارزمية HITS بجمع مجموعة وثائق أساســـية لاستعلام محدد مســـبقًا، ثم تقوم بشكل متكرر بحساب كل من المحور والسلطة لكل وثيقة. ولتجميع مجموعة الوثائق الأساسية، يتم استدعاء مجموعة الجذر التي تطابق الاســـتعلام من محرك بحث. ولكل وثيقة تم اســـترجاعها، تتم إضافة مجموعة من الوثائق التي تشـــير إلى الوثيقة الأصلية ومجموعة أخرى من الوثائق التي يشـــار اليها بواسطة الوثيقـــة الأصليـــة إلى المجموعة على أنها منطقة الوثيقة الأصلية. وهكذا تســـتمر عملية متكررة لتحديد هوية الوثيقة وتحليل الرابط حتى تتقارب قيم كل من المحور والسلطة، ثم يتم استخدام هذه القيم لفهرسة مجموعة الوثائق التي تم إنشاؤها لاستعلام محدد وترتيبها حسب الأولوية.

يعتـــبر التنقيب في هيكل الويب عملية اســـتخراج معلومات مفيدة مـــن الروابط المضمَّنة في وثائـــق الويب، ويتم اســـتخدامه لتحديد الصفحات الموثوقة والمحـــاور، والتي تعتبر حجر الزاوية في الخوارزميـــات المعاصرة للصفحات الأساســـية، والتي تعتبر محورية لمحركات البحث الشـــهيرة مثـــل Google وYahoo. ومثلما تشـــير الروابط الموضوعة لصفحة ويب إلى شـــعبية الموقع (أو موثوقيتـــه)، قد تشـــير الروابط الموجودة في صفحـــة الويب (أو موقع الويـــب الكامل) إلى عمق التغطيـــة لموضوع محدد. يعتـــبر تحليل الروابط مهمًا للغاية في فهم العلاقات المتبادلة بين أعداد كبيرة من صفحات الويب، مما يؤدي إلى فهم أفضل لمجتمع ويب معين.

أسئلة للمراجعة على قسم ٥-٧:

٦- ما هي بعض التحديات الرئيسة التي يفرضها الويب على اكتشاف المعرفة؟

٧- ما هو التنقيب في الويب؟ وكيف يختلف عن التنقيب المنتظم في البيانات أو التنقيب في النصوص؟

٨- ما هي المجالات الرئيسة الثلاثة للتنقيب في شبكة الإنترنت؟

٩- ما هو التنقيب في محتوى الويب؟ وكيف يمكن استخدامه لميزة تنافسية؟

١٠- ما هو التنقيب في هيكل (بنية) الويب؟ وكيف يختلف عن التنقيب في محتوى الويب؟

٨-٥ محركات البحث:

في الوقت الحالي، لا يستطيع أحد إنكار أهمية محركات البحث على الإنترنت. وتجدر الإشارة إلى أنــه كلما زاد حجم وتعقيد شــبكة الويب العالمية، أصبح العثور عــلى ما تريد عملية معقدة وشــاقة. ومن المعلوم أن محركات البحث تُستخدم لأسباب متنوعة. فيمكن استخدامها للتعرُّف عــلى منتج أو خدمة مــا قبل اتخاذ القرار بالــشراء (بما في ذلك اختيار البائــع أو مقدم الخدمة، والأســعار على المواقع / الباعة المتنوعة، والقضايا الشــائعة التي يناقشها الناس حول هذا المنتج، ومدى رضا المشــترين السابقين، وما هي المنتجات أو الخدمات الأخرى التي قد تكون أفضل، وما إلى ذلــك)، وللبحث عن الأماكن التي يجب زيارتها، والأشــخاص الواجب الالتقاء بهم، والأشــياء الواجــب القيام بها. بمعنى آخر، أصبحت محركات البحث محور معظم المعاملات المســتندة إلى الإنترنت وغيرها من الأنشــطة. ويستند هذا الادعاء إلى النجاح المذهل والشعبية الجارفة لشركة Google، وهي الشركة الأكثر انتشارًا في مجال محركات البحث. وما هو غامض إلى حدٍ ما بالنسبة للكثيرين هو كيف يقوم محرك البحث بتنفيذ ما يقصده المســتخدم. بعبارة أبسط، يعتبر محرك البحــث برنامجًا يقوم بالبحث عــن الوثائق (مواقع الإنترنت أو الملفات)، اســتنادًا على الكلمات المفتاحيــة التي يقدمها المســتخدمون، والتي تتعلق بموضوع بحثهم (ســواء كانت كلمات فردية أو مصطلحــات متعــددة الكلمات أو جملة كاملة). إن محركات البحث اليوم هي منبر الإنترنت، وبإمكانها الاستجابة لملايين من الاستفسارات بمئات اللغات المختلفة كل يوم.

من الناحية التقنية، فإن "محرك البحث" هو المصطلح الشــائع لأنظمة اســترداد المعلومات. وعلى الرغم من أن محركات البحث على الويب هي الأكثر شــيوعًا، فغالبًا ما تُستخدم في سياقات أخرى غير الويب، مثل محركات بحث سطح المكتب ومحركات بحث الوثائق. وكما سترى في هذا القســم، فإن العديد من المفاهيم والتقنيات التــي قمنا بتغطيتها في تحليلات النصوص والتنقيب في النــص في وقــت مبكر من هذا الفصــل تنطبق هنا أيضًا. ويعتبر الهــدف العام لمحرك البحث هو إرجاع وثيقة / صفحة واحدة أو أكثر والتي تتطابق بشــكل أفضل مع طلب بحث المستخدم، وفي حالــة وجود أكثر من وثيقة / صفحة واحدة، فغالبًا ما يتم عرضهم في قائمة مُرتَّبة وفقًا لمعيار محدد. وغالبًا ما يتم استخدام مقياسين لتقييم محركات البحث، وهما:

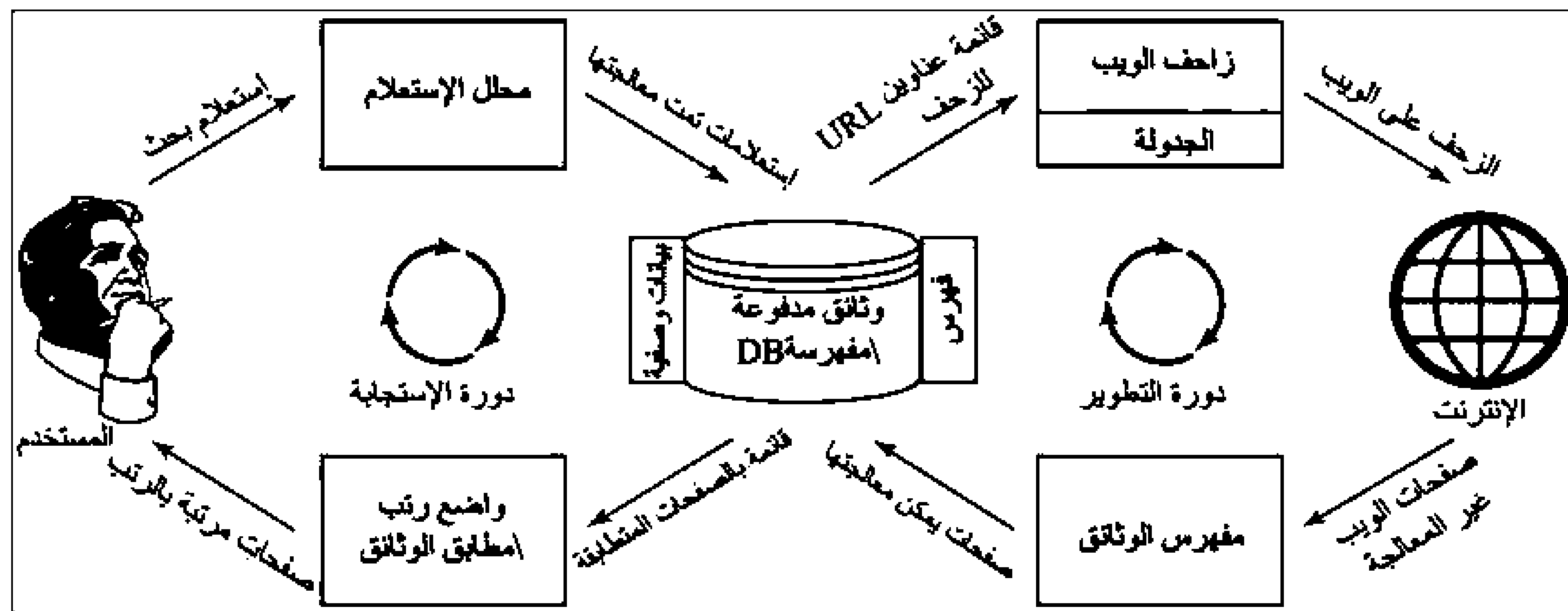
١- الفعالية (أو الجودة): وتعني العثور على الوثائق / الصفحات الصحيحة.

٢- الكفاءة (أو السرعة): وتعني إعادة الاستجابة بسرعة.

ويعمل هذان المقياســان في اتجاهين متعاكســين، بمعنى أن تحســين أحدهما يأتي على حساب الآخر، ويكون ذلك بالطبع وفقًا لتوقعات المستخدم. وأفضل محركات البحث هي تلك التي تتفوق في كل من المقياســين في نفس الوقت. ونظرًا لأن محركات البحث لا تبحث فقط، بل هي في الواقع تجد الوثائق / الصفحات وتعرضها، فربما كان هناك اسم أكثر ملاءمة لها وهو محركات الإيجاد.

تشريح محرك بحث:

الآن دعونا نحلل محرك بحث وننظر في داخله. على أعلى مســتوى، يتألف نظام محرك البحث من دورتين رئيســتين: دورة تطوير ودورة اســتجابة (ويعرض الشــكل (٥-١٢) هيكل محرك بحث نموذجــي على الإنترنت). ويمكننا تشــبيه دورة التطوير بعملية الإنتاج (تصنيــع وتخزين الوثائق / الصفحــات) ودورة الاســتجابة بعملية البيــع بالتجزئة (تزويد العملاء / المســتخدمين بما يريدون). وسيتم شرح هاتين الدورتين بمزيد من التفصيل في القسم التالي من هذا الفصل.



شكل ٥-١٢: هيكل نموذجي لمحرك بحث الإنترنت

١- دورة التطوير:

يعتــبر المكونان الرئيســان لدورة التطوير هما زاحف الويــب Web crawler ومفهرس الوثيقة document indexer، ويعد الغرض من هذه الدورة هو إنشــاء قاعدة بيانات ضخمة من الوثائق / الصفحات منظمة ومفهرســة بناءً على قيمة محتواها ومعلوماتها. ومما يظهر بوضوح شــديد أن الســبب في تطوير مثل هذا المستودع من الوثائق / الصفحات هو حجمه الكبير وتعقيده، فالبحث

في الويب للعثور على صفحات استجابة لتلبية احتياجات المستخدمين ليس عمليًا (أو ممكنًا في إطار زمني معقول)؛ ولذلك تقوم محركات البحث "بتخزين الويب" في قاعدة البيانات الخاصة بها، ومن ثم تستخدم النسخة المخبأة للويب للبحث عن وإيجاد كل ما يطلبه المستخدم. وبمجرد إنشائها، تسمح قاعدة البيانات هذه لمحركات البحث بالرد بسرعة وبدقة على استعلامات المستخدمين.

- زاحف الويب Web crawler: (ويسمى أيضًا العنكبوت أو عنكبوت الويب) وهو جزء من برنامج يقوم بالتصفح بشكل منهجي (أي يزحف عبر الإنترنت) بغرض العثور على صفحات الويب وجلبها. وفي كثير من الأحيان تقوم برامج زحف الويب بنسخ كل الصفحات التي تتم زيارتها وذلك لمعالجتها لاحقًا بواسطة وظائف أخرى لمحرك البحث. يبدأ زاحف الويب بإعداد قائمة من عناوين URL المراد زيارتها، ويتم سردها في جدول، وغالبًا ما يطلق عليها البذور. وقد تأتي عناوين URL هذه من تقديمات مشرفي المواقع أو - في أغلب الأحيان - تأتي من الروابط التشعبية الداخلية للوثائق / للصفحات التي تم الزحف إليها سابقًا. وبمجرد أن يزور الزاحف عناوين URL هذه، فإنه يحدد جميع الروابط التشعبية الموجودة في الصفحة ويضيفها إلى قائمة عناوين URL المراد زيارتها. وتتم زيارة عناوين URL المجدولة بشكل متكرر وفقًا لمجموعة من السياسات التي يحددها محرك البحث المحدد. ونظرًا لوجود عدد كبير جدًا من صفحات الويب، فيمكن للزاحف تنزيل عدد محدود فقط منها في وقت معين فقط؛ ولذلك فهي تحتاج إلى تحديد أولويات التنزيلات الخاصة بها.
- مفهرس الوثيقة document indexer: عندما يتم العثور على الوثائق وجلبها من خلال الزاحف، يتم تخزينها في منطقة مؤقتة لمفهرس الوثيقة وذلك لاستخراجها ومعالجتها. ويكون مفهرس الوثيقة مسؤولًا عن معالجة الوثائق (صفحات الويب أو ملفات الوثائق) ووضعها في قاعدة بيانات خاصة بها. ولتحويل الوثائق / الصفحات إلى التنسيق المرغوب فيه والقابل للبحث عنه بسهولة، يقوم مفهرس الوثائق بتنفيذ المهام التالية:
  - الخطوة ١- المعالجة المسبقة للوثائق: نظرًا لاحتمالية ظهور الوثائق المتحصل عليها عن طريق الزاحف في تنسيقات مختلفة، ولسهولة معالجتها أكثر، فسيتم في هذه الخطوة تحويلها كلها إلى أحد أنواع التمثيل القياسي. فمثلًا يتم فصل المحتويات المختلفة عن بعضها البعض ومن ثم تنسيقها (إذا لزم الأمر)، وتخزينها في مكان ما لمزيد من المعالجة. وتتضمن هذه المحتويات مثلًا (النصوص، الروابط التشعبية، الصور، إلخ).
  - الخطوة ٢- تحليل الوثائق: وهذه الخطوة هي أساسًا تطبيق أدوات وتقنيات التنقيب في النص (أي اللغويات الحاسوبية، ومعالجة اللغة الطبيعية NLP) إلى مجموعة من الوثائق / الصفحات. وفي هذه الخطوة، يتم أولًا تحليل الوثائق القياسية إلى مكونات لتحديد

الكلمات / المصطلحات الجديرة بالفهرســة، ثم بعد ذلك - وباستخدام مجموعة من القوانين - تتم فهرســة الكلمات / المصطلحات. وبشــكل أكثر تحديدًا - وباستخدام قواعد التحويل - يتم اســتخراج الكلمات / المصطلحات / الكيانات مــن الجمل الموجودة في هذه الوثائق، ثم باســتخدام معاجم خاصة، يتم تصحيح الأخطاء الإملائية والأخطاء الأخرى في هذه الكلمات / المصطلحات، مع ملاحظة عدم كون كل المصطلحات من أدوات التمييز. كما يتم اســتبعاد الكلــمات / المصطلحات غير التمييزية (والمعروفة أيضًا باســم كلــمات التوقف) من القائمة الجديرة بالفهرســة. ونظرًا لاحتمال ظهور نفس الكلمة / المصطلح بصيغ متعددة ومختلفة، فيتــم اتخاذ بعض الإجراءات التي تعمل على تقليل عدد الكلمات / المصطلحات ووضعها في صورها الأصلية. ومرةً أخرى، يتم اســتخدام المعاجم وموارد أخرى خاصة بلغة معينة (مثل WordNet)، كــما يتم تحديد المرادفات والمجانســات اللفظية، لمعالجة مجموعة الكلمات / المصطلحات قبل الانتقال إلى مرحلة الفهرسة.

- الخطــوة ٣- إنشــاء مصفوفــة "المصطلــح بحســب الوثيقــة" Term-by-document: في هذه الخطــوة، يتم تحديد العلاقات بــين الكلمات / المصطلحــات والوثائق / الصفحات. ويمكن أن يكون الوزن بســيطًا كتخصيص (١) للحضــور أو (٠) للغياب في الوثيقة / الصفحة. وعادةً ما يتم استخدام مخططات الوزن الأكثر تطورًا. وعلى سبيل المثال، في تعارض ثنائي، قد يتــم تخصيص التكرار كوزن (أي عدد مرات وجود نفس الكلمة / المصطلح في الوثيقة). وكما رأينا ســابقًا في هذا الفصل، فقد أشــارت الأبحاث والممارسات المتعلقة بالتنقيب عن النصوص بوضوح أن أفضل ترجيح قد يأتي من اســتخدام مصطلح التكرار مقســومًا على تكرار العكسي الوثيقة (TF / IDF). تقيس هذه الخوارزمية تكرار حدوث كل كلمة / مصطلح في وثيقة ما ثم تقارن ذلك التكرار في مقابل تكرار حدوثه في مجموعة الوثائق. وكما نعلم جميعًا، لا تعتبر كل الكلــمات / المصطلحــات ذات التكرار العالي عوامل تمييز جيــدة للوثيقة، وأيضًا فإن أحد العنــاصر الجيدة لتمييز الوثيقــة في مجال ما قد لا يكون جيدًا في مجال آخر. وبمجرد تحديد مخطط الوزن، يتم حساب الأوزان ويتم إنشاء ملف فهرس «المصطلح بحسب الوثيقة».

٢- دورة الاستجابة:

يعتبر المكونان الرئيســان لدورة الإســتجابة هما محلل الاستعلام query analyzer ومُطابق / مصنف الوثيقة document matcher/ranker.

- محلل الاســتعلام query analyzer: يكون محلل الاســتعلام مســؤولًا عــن تلقي طلب البحث من المســتخدم (عبر واجهــة خادم الويب لمحرك البحــث) وتحويله إلى هيــكل بيانات قياسي،

بحيث يمكن استجوابه / مطابقته بسهولة في مقابل الإدخالات في قاعدة بيانات الوثيقة. إن كيفية قيام محلل الاستعلام بما يفترض عمله تشابه تمامًا ما يقوم به مفهرس الوثيقة (كما تم توضيحه منذ قليل)؛ حيث يقوم محلل الاستعلام بتوزيع جملة البحث إلى كلمات / مصطلحات فردية باستخدام سلسلة من المهام التي تشمل التحويل، وإزالة كلمات التوقف، وتطبيق بعض الإجراءات، وتوضيح الكلمة / المصطلح (أي تحديد الأخطاء الإملائية، والمترادفات والمجانسات اللفظية). إن التشابه الوثيق بين محلل الاستعلام ومفهرس الوثيقة ليس من قبيل الصدفة. في الواقع، إن هذا التشابه منطقي تمامًا نظرًا لأن كليهما يعمل خارج قاعدة بيانات الوثائق؛ فأحدها يضع الوثائق / الصفحات باستخدام هيكل فهرس محدد، والآخر يقوم بتحويل سلسلة استعلام إلى نفس الهيكل بحيث يمكن استخدامها لتحديد موقع الوثائق / الصفحات ذات الصلة بسرعة.

- مُطابق / مصنف الوثيقة document matcher/ranker: وهنا تتم مطابقة بيانات الاستعلام المهيكلة مع قاعدة بيانات الوثيقة للعثور على الوثائق / الصفحات الأكثر ملاءمة وترتيبها أيضًا حسب ترتيب وثاقة الصلة بالموضوع / الأهمية. وربما يكون إتقان هذه الخطوة هو المكون الأكثر أهمية عند مقارنة محركات البحث المختلفة مع بعضها البعض. ويمتلك كل محرك بحث خوارزمية خاصة به (وغالبًا تكون ملكية مرخصة) يستخدمها لتنفيذ هذه الخطوة المهمة.

استخدمت محركات البحث في بداية ظهورها تطابقًا بسيطًا للكلمة المفتاحية مع قاعدة بيانات الوثائق وأعادت قائمة الوثائق / الصفحات المرتبة، حيث كان يعتبر تحديد الطلب هو إحدى الوظائف التي استخدمت عدد الكلمات / المصطلحات المتطابقة بين الاستعلام والوثيقة إلى جانب أوزان تلك الكلمات / المصطلحات. لم تكن جودة وفائدة نتائج البحث كلها جيدة. بعد ذلك، وتحديدًا في عام ١٩٩٧، توصل منشئو Google إلى خوارزمية جديدة تسمى Page Rank. وكما يشير الاسم، يعتبر Page Rank طريقة حسابية لترتيب الوثائق / الصفحات بناءً على مدى ملاءمتها وقيمتها / أهميتها. وعلى الرغم من كون Page Rank طريقة مبتكرة لتصنيف الوثائق / الصفحات، إلا أنه يعد زيادةً في عملية استرجاع الوثائق ذات الصلة من قاعدة البيانات وترتيبها بناءً على أوزان الكلمات / المصطلحات. تقوم Google بكل ذلك بشكل جماعي بغرض إعداد أفضل قائمة وثائق / صفحات من حيث الصلة بطلب البحث المعطَى. وبمجرد إنشاء قائمة مرتبة من الوثائق / الصفحات، يتم إعادتها مرةً أخرى إلى المستخدم بتنسيق سهل الاستيعاب. في هذه المرحلة، قد يختار المستخدم أي وثيقة من الوثائق الموجودة في القائمة والنقر عليها، وقد لا تكون هذه الوثيقة هي الأولى في أعلى القائمة. وإذا حدث هذا وتم اختيار وثيقة / صفحة غير الموجود في أعلى القائمة، فهل يمكننا أن نفترض أن محرك البحث لم يقم بترتيبهم ترتيبًا جيدًا؟ قد تكون

الإجابة نعم، حيث تقوم محركات البحث الرائدة مثل Google بمراقبة أداء نتائج البحث الخاصة بها من خلال التقاط وتسجيل وتحليل إجراءات وتجارب ما بعد تسليم المستخدمين. وغالبًا ما تؤدي هذه التحليلات إلى المزيد والمزيد من القواعد بهدف زيادة تحسين ترتيب الوثائق / الصفحات بحيث تكون الروابط في الأعلى هي الأكثر تفضيلًا للمستخدمين النهائيين.

تحسين محرك البحث:

إن تحسين محرك البحث (SEO) هي ممارسة يقصد منها التأثير على رؤية موقع التجارة الإلكترونية أو موقع الويب في نتائج البحث الطبيعية (غير المدفوعة أو الأساسية) لمحرك البحث. وبشكل عام، كلما ارتفعت الرتبة في صفحة نتائج البحث، وكلما تكرر ظهور الموقع في قائمة نتائج البحث، ازداد عدد الزوار الذين سيحصلون عليه من مستخدمي محرك البحث. وكإحدى إستراتيجيات التسويق عبر الإنترنت، تضع SEO في اعتبارها كيفية عمل محركات البحث، وما يبحث عنه الأشخاص، والمصطلحات الحقيقية للبحث أو الكلمات المفتاحية التي تتم كتابتها في محركات البحث، ومحركات البحث التي يفضلها الجمهور المستهدف. وقد ينطوي تحسين موقع الويب على تعديل محتواه وHTML والتشفير المرتبط به لزيادة مدى ملاءمته لكلمات رئيسة محددة ولإزالة الحواجز أمام أنشطة الفهرسة لمحركات البحث. كما يعد الترويج لموقع ما بغرض زيادة عدد الروابط الخلفية أو الداخلية، أحد الأساليب الأخرى لتحسين محركات البحث.

في السابق، ولكي تتم الفهرسة، كان على كل مشرفي المواقع إرسال عنوان صفحة، أو عنوان URL، إلى مختلف المحركات، والتي ستقوم بعد ذلك بإرسال "عنكبوت" لكي يقوم بـ "الزحف" تجاه تلك الصفحة، ومن ثم يستخرج منها الروابط إلى صفحات، ثم يقوم بإرجاع المعلومات التي تم العثور عليها على الصفحة إلى الخادم لتتم الفهرسة. وكما تم التوضيح من قبل، تتضمن هذه العملية قيام عنكبوت محرك بحث بتنزيل صفحة وتخزينها على الخادم الخاص بمحرك البحث، حيث يقوم برنامج آخر - يعرف باسم المفهرس - باستخلاص معلومات متنوعة عن تلك الصفحة، مثل الكلمات التي تحتويها ومواقع تلك الكلمات بالتحديد، بالإضافة إلى أي وزن لكلمات محددة، وكل الروابط التي تحتوي عليها الصفحة، والتي تتم جدولتها بعد ذلك بغرض القيام بالزحف في وقت لاحق. أما الآن، فلم تعد محركات البحث تعتمد على تقديم مشرفي المواقع عناوين URL (على الرغم من قدرتهم على فعل ذلك حتى الآن)؛ ولكنهم يمارسون الزحف الحثيث والمستمر على الويب لإيجاد وجلب وفهرسة كل شيء يتعلق بالبحث.

إن إجراء الفهرسة بواسطة محركات بحث مثل Google وBing وYahoo ليس جيدًا بما فيه الكفاية للأنشطة التجارية، حيث إن اكتساب ترتيب على محركات البحث الأكثر استخدامًا

والحصــول على مرتبة أعلى من منافســيك هو ما يصنــع الفرق (انظر الفكــرة التطبيقية (٥-٣) للحصــول على قائمة محركات البحث الأكثر اســتخدامًا). ويمكن رفــع ترتيب صفحة ويب داخل نتائج البحث باســتخدام مجموعة متنوعة من الأســاليب. كما أن الارتباط المتبادل بين صفحات نفس موقع الويب يوفر مزيدًا من الروابط إلى الصفحات الأكثر أهمية مما يزيد من فرص رؤيتها ضمن نتائج البحث. وســتؤدي كتابة المحتوى الذي يتضمن عبارات الكلمات المفتاحية التي يتم البحث عنها بشــكل متكرر - بحيث تكون ذات صلة بمجموعة كبيرة من استعلامات البحث - إلى زيــادة عدد الزيارات. كما أن تحديــث المحتوى بغرض الحفاظ على الزحف إلى محركات البحث بشكل متكرر يضفي على الموقع زيادة في الوزن. وكذلك تؤدي إضافة كلمات رئيسة ذات صلة إلى البيانات الوصفية لصفحة الويب - بما في ذلك علامة العنوان والوصف الدقيق - إلى تحسين مدى ملاءمــة قوائم البحث الخاصة بالموقع، وبالتالي زيادة عدد الزيارات. ويمكن أن تســاعد تســوية عناوين URL لصفحات الويب - بحيث يمكن الوصول إليها عبر عناوين URL متعددة واستخدام عناصر ربط العناوين الأساسية والاتجاهات - في التأكد من أن جميع الروابط المؤدية إلى إصدارات مختلفة من عنوان URL يتم احتسابها في جميع أنحاء نقاط انتشار رابط الصفحة.

رؤية فنية ٥-٣:

أعلى ١٥ محرك بحث الأكثر انتشارًا (أغسطس ٢٠١٦)

نعــرض فيــما يلي أعلى ١٥ محــركًا من محــركات البحث الأكثر انتشــارًا ومصدرها مــن: (ebizMBA Rank ذكاء الأعــمال zmba.com/articles/search-ngines)، وهو متوسط يتم تحديثه باســتمرار لكل من موقع Alexa Global Traffic Rank، و U.S.Traffic Rank من كل من Compete وQuantcast.

الترتيب	محرك البحث	التقدير الشهري لعدد الزوار
١	Google	١٦٠٠٠٠٠٠٠٠
٢	bing	٤٠٠٠٠٠٠٠٠
٣	Yahoo Search	٣٠٠٠٠٠٠٠٠
٤	Ask	٢٤٥٠٠٠٠٠٠
٥	AOL Search	١٢٥٠٠٠٠٠٠
٦	Wow	١٠٠٠٠٠٠٠٠
٧	WebCrawler	٦٥٠٠٠٠٠٠

الترتيب	محرك البحث	التقدير الشهري لعدد الزوار
٨	MyWebSearch	٦٠٠٠٠٠٠٠
٩	Infospace	٢٤٠٠٠٠٠٠
١٠	Info	١٣٥٠٠٠٠٠
١١	DuckDuckGo	١١٠٠٠٠٠٠
١٢	Contenko	١٠٥٠٠٠٠٠
١٣	Dogpile	٧٥٠٠٠٠٠
١٤	Alhea	٤٠٠٠٠٠٠
١٥	ixQuick	١٠٠٠٠٠٠

طرق تحسين محركات البحث:

بشكل عام، يمكن تصنيف تقنيات تحسين محركات البحث إلى فئتين رئيستين:

١- التقنيات التي توصي بها محركات البحث كجزء من التصميم الجيد للموقع.

٢- التقنيات التي لا تقبلها محركات البحث.

تحاول محركات البحث تقليل تأثير النوع الثاني - والذي غالبًا ما يسمى spamdexing (ويُعرف أيضًا باسم البحث غير المرغوب فيه search spam، أو محرك بحث غير مرغوب فيه search engine spam، أو تسمم محرك البحث search engine poisoning). وقد تم تصنيف هذه الأساليب والممارسين لها إلى قسمين: أدوات SEO ذات قبعة بيضاء وأدوات SEO ذات قبعة سوداء (Goodman, 2005)؛ حيث تميل القبعات البيضاء إلى إفراز نتائج تدوم لفترة طويلة، في حين تتوقع القبعات السوداء حظر مواقعها في نهاية المطاف إما بشكل مؤقت أو دائم بمجرد أن تكتشف محركات البحث ما تفعله.

تعتبر تقنية تحسين محركات البحث (SEO) قبعة بيضاء إذا كانت متوافقة مع إرشادات محرك البحث ولا تحتوي على أي خداع. ونظرًا لأن إرشادات محرك البحث لا تُكتب كسلسلة من القواعد أو الوصايا، فإن هذا تمييز مهم يجب ملاحظته. ولا يقتصر دور SEO ذات القبعة البيضاء على اتباع الإرشادات فحسب، بل يمتد إلى ضمان أن المحتوى الذي يفهرسه محرك البحث ثم يصنف لاحقًا هو

نفس المحتوى الذي سيراه أي مستخدم. وعادةً ما يتم تلخيص نصيحة القبعة البيضاء على أنها إنشاء محتوى للمستخدمين، وليس لمحركات البحث، ثم تسهيل الوصول لهذا المحتوى بواسطة العناكب، بدلًا من محاولة خداع الخوارزمية من الغرض المراد منها. كما يعتبر SEO ذو القبعة البيضاء أحد الأوجه العديدة المشابهة لتطوير الويب الذي يدعم إمكانية الوصول، على الرغم من تطابقهما.

يحاول SEO ذو القبعة السوداء تحسين الترتيب بطرق غير مقبولة من جانب محركات البحث، أو بأساليب تنطوي على شيء من الخداع. تستخدم تقنية القبعة السوداء النص المخفي، إما بتلوينه بنفس لون الخلفية أو بعلامة "div" غير مرئية أو بوضعه خارج الشاشة. وتعطي طريقة أخرى صفحة مختلفة اعتمادًا على كون الصفحة مطلوبة سواء بواسطة زائر بشري أو محرك بحث، وهي تقنية تُعرف باسم إخفاء الهوية أو التغطية cloaking. كما قد تعمد محركات البحث إلى معاقبة المواقع التي تكتشفها باستخدام أساليب القبعة السوداء، إما عن طريق تقليل ترتيبها أو إزالة قوائمها من قواعد بياناتها تمامًا. ويمكن توقيع هذه العقوبات إما تلقائيًا من خلال خوارزميات محركات البحث أو من خلال مراجعة يدوية للموقع. ومن الأمثلة على ذلك، ما قامت به شركة Google في فبراير ٢٠٠٦ عندما قامت بإزالة كل من BMW ألمانيا وRicoh ألمانيا لقيامهما بممارسات غير معتمدة (Cutts, 2006). ومع ذلك، سارعت الشركتان بتقديم الاعتذار وتصحيح ممارساتهما، ومن ثم تمت إعادتهما إلى قائمة Google.

بالنسبة لبعض الأنشطة التجارية، قد يحقق SEO عائدًا كبيرًا على الاستثمار. ومع ذلك، يجب الأخذ في الاعتبار أن محركات البحث غير مدفوعة لحركة البحث المجاني، وأن خوارزمياتها تتغير باستمرار، ولا توجد ضمانات لاستمرار الإحالات. وبسبب هذه الحالة من عدم اليقين أو عدم الاستقرار، فقد تتكبد الأعمال التجارية التي تعتمد بشكل كبير على زيارات محركات البحث خسائر كبيرة إذا قرر محرك البحث تغيير خوارزمياته والتوقف عن استقبال الزوار. يقول Eric Schmidt الرئيس التنفيذي لشركة Google: "أجرت Google في عام ٢٠١٠ أكثر من ٥٠٠ تغيير خوارزمية، أي ما يقرب من ١٫٥ تغيير في اليوم الواحد". ونظرًا لصعوبة مواكبة قواعد محرك البحث المتغيرة باستمرار، فإن الشركات التي تعتمد على حركة البحث تمارس واحدًا أو أكثر من الإجراءات التالية:

١- استئجار شركة متخصصة في تحسين محركات البحث (والتي يتوافر عدد كبير منها في هذه الأيام) وذلك لتحسين جاذبية الموقع الخاص بها باستمرار لتغيير ممارسات محركات البحث.

٢- تشجيع شركات محركات البحث لإدارج أسمائهم في أقسام الجهات الراعية المدفوعة.

٣- النظر في التحرر من الاعتماد على حركة محرك البحث.

إن أهــم مــا يميز أي مواقع تجــارة إلكترونية - ســواء كانت معتمدة على محــركات البحث (بعضوية أو بدون) أو قادمة من مواقع وأماكن أخرى - هو زيادة احتمالية معاملات العملاء إلى أقصى حد. ولا يعتبر وجود عدد كبير من الزائرين دون مبيعات هو ما تم تصميمه لموقع التجارة الإلكترونية النموذجي. وتتعلق الحالة العملية (٥-٧) بمركز تسوق كبير يعتمد على الإنترنت حيث يتم استخدام تحليل مفصل لسلوك العميل (باستخدام نقرات ومصادر البيانات الأخرى) لتحسين معدل التحويل بشكل ملحوظ.

حالة عملية ٥-٧

فهْم لماذا يؤدي التنازل عن عربات التسوق إلى زيادة مبيعات بمقدار ١٠ مليون دولار

قــام مركــز Lotte.com، وهو مركز تســوق الإنترنت الرائد في كوريــا والذي يضم ١٣ مليــون عميل، بتطوير نظام متكامل لتحليل حركة زيارات الويب باســتخدام SAS لتحليلات تجربة العملاء. ونتيجة لذلك، تمكن Lotte.com من تحسين تجربة الإنترنت لعملائها، فضلًا عن تحقيق عوائد أفضل من حملاتها التسويقية. والآن، يمكن للمديرين التنفيذيــين لـ Lotte.com تأكيد النتائج في أي مــكان وفي أي وقت، بالإضافة إلى إجراء تغييرات فورية.

ومع ما يقرب من مليون زائر لموقع الويب كل يوم، كان Lotte.com بحاجة إلى معرفة عدد الزوار الذين يقومون بالشراء وما هي القنوات التي تجلب الزيارات الأكثر قيمة. بعد استعراض العديد من الحلول والأســاليب المتنوعة، قدم Lotte.com نظامًا متكاملًا خاصًا به لتحليل حركة زيارات الويب باستخدام حل SAS لتحليل تجارب العملاء، وهذا هو أول نظام للتحليل السلوكي على الإنترنت يطبق في كوريا.

وباســتخدام هذا النظام، يمكن لمركز Lotte.com قياس وتحليل كل من أعداد زوار موقع الويب بدقة، والحالة الراهنة لعرض الصفحة أمام زوار الموقع ومبرمي الصفقات، وشــعبية كل فئة من فئات المنتجات وكذلك شــعبية المنتجات نفســها، والنقرات على التفضيــلات الخاصة بــكل صفحة، وفعاليــة الحملات، وغير ذلك الكثــير. أي أن هذه المعلومات تمكّن Lotte.com من فهم العملاء وســلوكهم على الإنترنت بشــكل أفضل، وإجراء التسويق المستهدف المتطور والفعال من حيث التكلفة.

وتعليقًــا عــلى هذا النظــام، قال Jung Hyo-hoon مســاعد المديــر العام بفريق تخطيط التســويق لموقع Lotte.com: "نتيجةً لإدخال نظام تحليل SAS، تم اكتشــاف العديد من"الحقائق الجديدة"حول سلوك العملاء، وبعضها كانت «حقائق غير مريحة»

وأضاف: "في الواقع كان لبعض أنشطة تخطيط الموقع التي تم الاضطلاع بها مع توقع نتائج معينة رد فعل ضعيف من جانب العملاء، واحتاج مخططو الموقع وقتًا طويلًا للتعرف على هذه النتائج".

فوائد:

وقد أدى تقديم حلول SAS لتحليل تجربة العملاء إلى تحويل موقع الويب Lotte.com بشكل كامل. ونتيجة لذلك، تمكن Lotte.com من تحسين تجربة الإنترنت لعملائه بالإضافة إلى تحقيق عوائد أفضل من حملاته التسويقية. ومنذ تطبيق SAS لتحليلات تجربة العملاء، جنى Lotte.com العديد من الفوائد، منها:

قفزة في ولاء العملاء:

يمكن جمع كمية كبيرة من معلومات النشاط المتطورة في إطار بيئة الزائر، بما في ذلك جودة الزيارات. وقال Jung: "إنه من خلال تحليل الزيارات الصالحة الفعلية والبحث في صفحة أو صفحتين فقط، يمكننا تنظيم حملات لرفع مستوى الولاء، ومن ثم تحديد نطاق معين من التأثير. وبالإضافة لذلك، من الممكن تصنيف وتأكيد معدل الطلب لكل قناة ومعرفة أي القنوات تحتوي على أكبر عدد من الزائرين".

تحليل كفاءة التسويق الأمثل:

لا يقتصر النظام على مجرد تحليل أرقام الزائرين فقط، بل يستطيع تحليل معدل التحويل (عربة التسوق، الشراء الفوري، قائمة الرغبات، إتمام الصفقة) مقارنةً بالزوار الفعليين لكل نوع من أنواع الحملات (الانتساب أو البريد الإلكتروني، اللافتات، الكلمات المفتاحية، وغيرها)، لذلك يمكنه أيضًا إجراء تحليل مفصل لفعالية القناة. بالإضافة إلى ذلك، يمكنه التأكد من كلمات البحث الأكثر استخدامًا من جانب الزوار لكل نوع حملة والموقع والمنتجات المشتراة. كما يمكنه عن طريق دالة "the page overlay" قياس عدد النقرات وعدد الزوار لكل عنصر في أي صفحة لقياس قيمة كل موقع في الصفحة. تمكن هذه الإمكانية Lotte.com من الاستبدال أو التجديد الفوري للعناصر ذات الزيارات القليلة.

رضا العملاء المحسّن وتجاربهم يؤديان إلى ارتفاع المبيعات:

قام Lotte.com بإنشاء قاعدة بيانات تحليل سلوك العملاء التي تقيس كل زائر، والصفحات التي تتم زيارتها، وكيف يتنقل الزوار في الموقع، والأنشطة التي يتم تنفيذها لتمكين التحليل المتنوع وتحسين كفاءة الموقع. بالإضافة إلى ذلك، تحتفظ قاعدة

البيانـــات بمعلومات عن التوزيعات الســكانية للعملاء وحجم عربة التســوق ومعدل التحويل وعدد الطلبات وعدد المحاولات.

وبتحليل مراحل عمليــة الطلب وتحديد المرحلة التي تعيق معظم العملاء ومن ثم تثبيــت هذه المراحل، يمكن زيادة معدلات التحويل. في الســابق، تم التحليل فقط على الطلبــات الموضوعــة. ومن خلال تحليل نمــط حركة الزوار قبل الطلــب، وعند النقطة التي يحدث فيها الانفصال، يمكن توقع ســلوك العملاء كما يمكن القيام بأنشطة تسويق متطــورة. ومن خلال تحليل نمــط الزائرين، يمكن أن تتأثر عمليات الشراء بشــكل أكثر فاعليــة وبالتــالي ينعكس طلب العملاء بشــكل فوري لضمان الحصول على اســتجابات أسرع. وقــد حدث تحســن في رضا العمــلاء لوجود فكرة أفضل لــدى Lotte.com عن سلوكيات واحتياجات واهتمامات كل عميل.

وفي تقييمــه للنظام، علق Jung قائلًا: "يمكن تحديد تحســينات في خدمة العملاء ومواضيع التسويق المســتهدفة من خلال معرفة كيفية تحرك كل مجموعة من العملاء على أساس البيانات،مما ساعد على نجاح عدد من الحملات".

ومــع ذلك، فإن الفائدة الأكثر أهمية من هذا النظام هي اكتســاب نظرة ثاقبة عن العملاء الأفراد ومجموعــات العملاء المختلفة. ومن خلال فهم التوقيت الذي يبرم فيه العمــلاء صفقات الشراء وأســلوبهم في التنقــل عبر صفحة الويب، يمكــن الآن تحقيق التسويق المستهدف للقنوات وتحسين تجربة العملاء.

بالإضافة إلى ذلك، فقد أدى تنفيذ SAS لتحليلات تجربة العملاء من قِبل أكبر موزع عــبر البحار لمركز Lotte.com إلى زيادة في مبيعات الســنة الأولى بقيمة ٨ ملايين يورو (أي ما يعادل ١٠ ملايين دولار أمريكي) من خلال تحديد أسباب هجر عربة التسوق.

أسئلة للمناقشة:

١- كيف استخدم Lotte.com التحليلات لتحسين المبيعات؟

٢- ما هي التحديات والحل المقترح والنتائج التي تم الحصول عليها؟

٣- هــل تعتقــد أن شركات التجارة الإلكترونية تعتبر في وضع أفضل للاســتفادة من فوائد التحليلات؟ ولماذا؟ وكيف؟

Sources: From Heightening customer loyalty and optimizing channels, SAS, Customer Success Stories.sas.com/success/lotte.html (accessed March 2013).

أسئلة مراجعة على القسم ٥-٨:

٥- ما هو محرك البحث؟ ولماذا تهتم الشركات اليوم بمحركات البحث؟

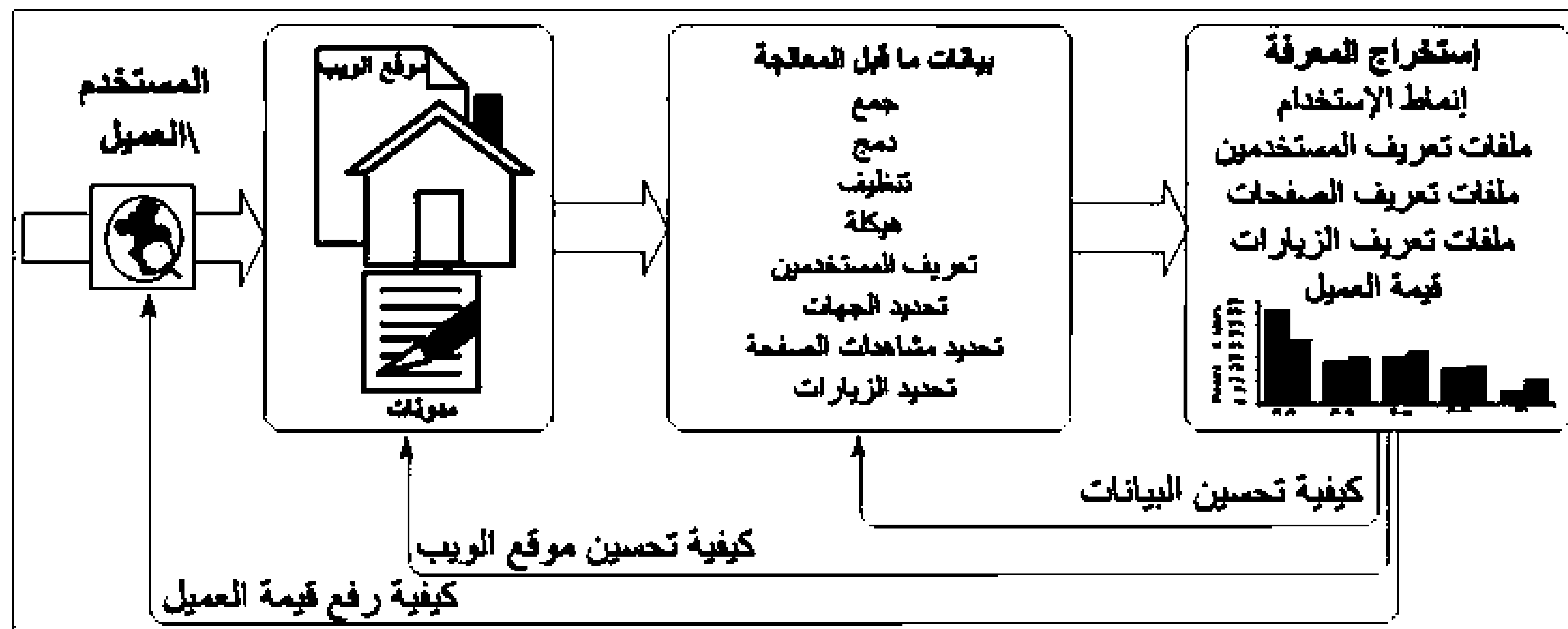
٦- ما هو زاحف الويب Web crawler؟ وفيم يُستخدم؟ وكيف يعمل؟

٧- ما هو "تحسين محرك البحث؟" ومن الذي يستفيد منه؟

٨- ما هي الأشياء التي يمكن أن تساعد صفحات الويب على احتلال ترتيب أعلى في نتائج محرك البحث؟

٥-٩ التنقيب في استخدام الويب (تحليلات الويب):

التنقيب في استخدام الويب (ويسمى أيضًا تحليلات الويب) هو استخراج معلومات مفيدة من البيانات التي يتم إنشاؤها من خلال زيارات صفحة الويب والمعاملات. ويمكن أن يساعدنا تحليل المعلومات التي يتم جمعها بواسطة خوادم الويب على فهم سلوك المستخدم بشكل أفضل. وغالبًا ما يطلق على تحليل هذه البيانات تحليل مسار النقر. وقد تتمكن الشركة من تمييز أنماط مثيرة للاهتمام من مسار النقر، وذلك من خلال استخدام تقنيات التنقيب في البيانات والنصوص. فعلى سبيل المثال، قد تعلم أن ٦٠٪ من الزائرين الذين بحثوا عن "فنادق في مأوى" قد بحثوا في وقت سابق عن "رحلات جوية إلى مأوى". هذه المعلومات قد تكون مفيدة في تحديد مكان وضع الإعلانات عبر الإنترنت. كما أن تحليل مسار النقر قد يكون مفيدًا أيضًا في معرفة وقت وصول الزائرين إلى الموقع. فمثلًا، إذا كانت إحدى الشركات تعرف أن ٧٠٪ من تنزيلات البرامج من موقعها على الويب قد حدثت بين الساعة السابعة والحادية عشرة مساءً، فإنها قد تخطط للوصول إلى دعم أفضل للعملاء وسعة تحميلية أعلى للشبكة أثناء تلك الفترة. ويعرض الشكل (٥-١٣) عملية استخراج المعرفة من بيانات مسار النقر وكيف يتم استخدام تلك المعرفة لتحسين كل من العملية وموقع الويب، والأكثر أهمية هو رفع قيمة العميل.



الشكل ٥-١٣: استخراج المعرفة من بيانات استخراج الويب

تقنيات تحليلات الويب:

هناك العديد من الأدوات والتقنيات لتحليلات الويب في السوق. وتزداد شعبية أدوات تحليل بيانات الإنترنت، نظرًا لقدرتها على قياس بيانات الإنترنت وجمعها وتحليلها لفهم الاستخدام الأمثل للويب وتحسـينه. وتبشر تحليلات الويب بإحداث ثورة في كيفية إنجاز الأعمال على الويب، كما يمكن استخدامها كأداة للأعمال التجارية الإلكترونية وأبحاث السوق ولتقييم وتحسين فعالية مواقع الويب الخاصة بالتجارة الإلكترونية. ويمكن لتطبيقات تحليلات الويب أيضًا مسـاعدة الشركات في قياس نتائج الحملات الإعلانية المطبوعة أو الإذاعية التقليدية. وبالإضافة إلى ذلك، فإنه يمكنها أن تسـاعد في تقدير كيفية تغير حركة المرور إلى موقع ويـب بعد إطلاق حملة إعلانية جديدة. كما توفر تحليلات الويب معلومات حول عدد الزائرين لموقع الويب وعدد مشاهدات الصفحة، فضلًا عن مساعدتها في قياس اتجاهات الحركة والرواج، والتي يمكن استخدامها في أبحاث السوق.

وهنـاك فئتان رئيسـتان لتحليلات الويب وهما: خارج الموقع، وفي الموقع. وتشـير تحليلات الويب خارج الموقع إلى قياس الويب والتحليلات عنك وعن منتجاتك التي تتم خارج موقع الويب الخاص بك، وهو يشمل قياس نسبة المتابعين المحتملين لموقع الويب (احتمال أو فرصة)، ومشاركة الصوت (الرؤية أو الكلام الشفهي)، والضجيج (التعليقات أو الآراء) الذي يحدث على الإنترنت.

وقد كانت تحليلات الويب في الموقع هي الأكثر شيوعًا. فمن الناحية التاريخية، أشارت تحليلات الويب إلى قياس نسبة زائري الموقع. ومع ذلك، فقد تضاءل هذا الأمر في السنوات الأخيرة، ويرجع ذلك بشـكل أسـاسي إلى أن الموردين ينتجون أدوات تمتد لتشـمل كل الطبقات. وتقيس تحليلات الويب في الموقع سـلوك الزائرين عندما يكونون على موقع الويب الخاص بك، ويشـمل ذلك برامج التشـغيل والتحويلات -مثل درجة ارتباط الصفحات المقصودة المختلفة، بالمشـتريات عبر الإنترنت. كما تقوم تحليلات الويب الموجودة في الموقع بقياس أداء موقع الويب الخاص بك في سياق تجاري، ثم تتم مقارنة البيانات المجمعة على موقع الويب بمؤشرات الأداء الرئيسـة للأداء واسـتخدامها في تحسين استجابة الجمهور لموقع الويب أو حملة التسويق. وعلى الرغم من أن Google Analytics هي خدمة تحليلات الويب الأكثر استخدامًا على نطاق واسع، إلا أن خدمة Yahoo ومايكروسوفت، والأدوات الجديدة والأفضل التي تظهر باستمرار، توفر طبقات إضافية من المعلومات.

وبالنسـبة إلى تحليـلات الويب في الموقع، فهنـاك طريقتان تقنيتان لجمـع البيانات. وتتمثل الطريقـة الأولى والأكثر تقليدية في تحليل ملف سـجل الخادم، حيث يقوم خادم الويب بتسـجيل طلبـات الملفـات التي تقدمها المتصفحات. وأمـا الطريقة الثانية فتتمثـل في وضع العلامات على الصفحات، والتي تسـتخدم JavaScript مضمنًا في شـفرة صفحة الموقع لتقديم طلبات الصور إلى

خـادم مخصص للتحليلات تابع لجهة خارجية كلما تم عـرض صفحة من خلال متصفح الويب (أو عنـد حدوث نقرة ماوس). ولإنتاج تقارير المرور على شـبكة الإنترنت، يمكن معالجة كل من جمع البيانات فضلًا عن هذين المسـارين الرئيسـين، كما يمكن أيضًا إضافة مصادر بيانات أخرى لزيادة بيانات سـلوك موقع الويب. وقد تتضمن هـذه المصادر الأخرى البريد الإلكتروني، أو بيانات حملة البريد المباشر، أو المبيعات، أو تاريخ القيادة، أو البيانات التي نشأت عبر وسائل التواصل الاجتماعي.

مقاييس تحليلات الويب:

توفر برامج تحليلات الويب الوصول إلى الكثير من البيانات التسويقية القيّمة، والتي يمكن الاستفادة منها للحصول على إحصاءات أفضل لتنمية أعمالك وتوثيق عائد الاستثمار (ROI) بشكل أفضل، وذلك باسـتخدام مجموعة متنوعـة من مصادر البيانات. ويمكن اسـتخدام البصيرة والذكاء المكتسـبة من تحليلات الويب لإدارة الجهود التسـويقية للمؤسسـة ومختلف منتجاتها أو خدماتها بفعالية. وتوفر برامج تحليلات الويب بيانات بشكل فوري تقريبًا، والتي يمكنها توثيق نجاحات حملاتك التسويقية، أو تمكينك من إجراء تعديلات في الوقت المناسب على إستراتيجيات التسويق الحالية.

وفي حـين أن تحليلات الويب توفر نطاقًا واسـعًا من المقاييس، فهنـاك أربع فئات من المقاييس قابلة للتطبيق بشكل عام ويمكن أن تؤثر بشكل مباشر على أهداف أعمالك (The Westover Group, 2013). وتشمل هذه الفئات ما يلي:

- سهولة استخدام موقع الويب: كيف كانوا يستخدمون الموقع الخاص بي على الويب؟
- مصادر الزيارات: من أين أتوا؟
- ملامح الزوار: كيف يبدو زوار موقعي؟
- إحصائيات التحويل: ماذا يعني كل هذا للأعمال؟

قابلية استخدام موقع الويب:

بدايـة من موقع الويب الخاص بك، دعنـا نلقي نظرة على مدى نجاحه مع زوارك. فهذا هو المكان الذي يمكنك من خلاله معرفة مدى «سـهولة اسـتخدامه» بالفعل أو ما إذا كان المحتوى الذي يقدمه مناسبًا أم لا.

١- استعراضات الصفحة: وهو أبسط القياسات، ويتم تقديم هذا المقياس عادةً على أنه "متوسط عدد مشاهدات الصفحة لكل زائر". فإذا جاء الأشخاص إلى موقعك على الويب ولم يشاهدوا العديد من الصفحات، فقد يواجه موقعك الإلكتروني مشـكلات في تصميمه أو بنيته. وهناك

تفســير آخر لانخفاض عدد مشاهدات الصفحة وهو قطع الرسائل التسويقية التي جلبتها إلى الموقع والمحتوى المتاح بالفعل.

٢- الوقت على الموقع: ويشــبه مرات مشــاهدة الصفحة، فهو مقياس أســاسي لتفاعل الزائر مع موقع الويب الخاص بك. وبشــكل عام، فكلما زاد إنفاق الشــخص على موقع الويب الخاص به، كان ذلك أفضل. وقد يعني ذلك أنهم يراجعون المحتوى الخاص بك بعناية، مســتخدمين المكونــات التفاعلية المتوفرة لديك، والبناء للوصول إلى قرار مســتنير للشراء أو الاســتجابة أو اتخــاذ الخطــوة التالية التي قدمتها. وعلى العكس من ذلك، فإنــه يجب أيضًا فحص الوقت المســتغرق في الموقع مقابل عدد الصفحات التي تم عرضها للتأكد من أن الزائر لا يقضي وقته في محاولة العثور على محتوى يسهل الوصول إليه.

٣- التنزيــلات: يتضمن ذلك ملفات PDF ومقاطع الفيديو والموارد الأخرى التي توفرها لزائريك. ولا بد أن تضع في اعتبارك مدى ســهولة الوصول إلى هــذه العناصر وكذلك مدى الترويج لها. فمثلًا، إذا كانت إحصائيات الويب الخاصة بك تكشــف أن ٦٠٪ من الأفراد الذين يشــاهدون فيديو تجريبي يقومون أيضًا بعملية شراء، فحينئذٍ ســترغب في وضع إستراتيجية لزيادة نسبة المشاهدة لهذا الفيديو.

٤- خريطة النقر: يمكن لمعظم برامج التحليلات عرض النســبة المئوية للنقرات التي حصل عليها كل عنــصر على صفحة الويب الخاصة بك. ويتضمن هذا صورًا قابلة للنقر، وروابط نصية في نسختك، وتنزيلات، وبالطبع أيضًا أي تنقل قد يكون لديك على الصفحة. هل يقومون بالنقر على العناصر الأكثر أهمية؟

٥- مســارات النقر: على الرغم من أن تقييم مســارات النقر ينطوي على مزيد من المشــاركة، إلا أنه يمكنه الكشــف بسرعة أين يمكن أن تفقد الزوار في عملية محددة. ويســتخدم موقع ويب المصمم بشكل جيد مزيجًا من بنية الرسومات والمعلومات لتشجيع الزائرين على اتباع المسارات "المحددة مســبقًا" من خلال موقع الويب الخاص بك. وهذه المســارات ليســت صلبة ولكنها خطوات بديهية تتوافق مع مختلف العمليات التي قمت بإنشائها في موقع الويب. وقد تكون إحــدى هذه العمليات هي "تعليم" الزائر الذي لديه أدنى فهم لمنتجك أو خدمتك. وقد تكون الأخــرى عملية "تحفيــز" للزائر الذي يعيد النظــر مرة أخرى في تحديث أو إعــادة شراء. أما العملية الثالثة فقد يتم تنظيمها حول العناصر التي تسوقها عبر الإنترنت. وسوف يكون لديك العديد من مسارات العمليات في موقع الويب الخاص بك كما يكون لديك الجماهير المستهدفة والمنتجات والخدمات، ويمكن قياس كل منها من خلال تحليلات الويب لتحديد مدى فعاليتها.

مصادر الزيارات:

برنامج تحليلات الويب الخاص بك هو أداة لا تصدق لتحديد مكان الزيارات الخاصة بك على الويب. ويتم تصنيف الفئات الأساسية مثل محركات البحث ومواقع الويب المرجعية والزيارات من الصفحات التي تم وضع إشارة عليها (مثل: المباشرة) مع القليل من مشاركة المسوق. كما يمكنك أيضًا تحديد زيارات الويب التي تم إنشاؤها بواسطة حملاتك الإعلانية المتعددة غير المتصلة أو عبر الإنترنت، بقليل من الجهد.

١- مواقع الويب التي تقوم بإحالة الزائرين إلى موقع الويب الخاص بك Referral: إن مواقع الويب التي تحتوي على روابط تقوم بإرسال الزائرين مباشرةً إلى موقع الويب الخاص بك، تعتبر مواقع ويب مرجعية، وسيقوم برنامج التحليلات الخاص بك بتحديد كل موقع يتم إرسال الزيارة لك منه، كما يوجد تحليل أعمق سيساعدك على تحديد الإحالات التي تنتج الحجم الأكبر، والتحويلات الأعلى، والزوار الأحدث، وما إلى ذلك.

٢- محركات البحث: يتم تقسيم البيانات في فئة محركات البحث بين البحث مدفوع الثمن والبحث الأساسي (أو الطبيعي). ويمكنك مراجعة أهم الكلمات المفتاحية التي أنتجت الزيارة إلى موقعك ومعرفة ما إذا كانت هذه الكلمات تمثل منتجاتك وخدماتك أم لا. واعتمادًا على عملك قد ترغب في الحصول على مئات (أو آلاف) الكلمات المفتاحية التي تحدد ملامح العملاء المحتملين. وحتى أبسط بحث عن المنتج يمكن أن يكون بمتغيرات تعتمد على شكل العبارات الفردية في طلب البحث.

٣- المباشرة: وتُنسب عمليات البحث المباشرة إلى مصدرين. يتمثل الأول في الشخص الذي يضع إشارة مرجعية على إحدى صفحات الويب الخاصة بك في صفحاته المفضلة ويضغط على هذا الرابط، فيتم تسجيله كبحث مباشر. أما المصدر الثاني فيتمثل في كتابة شخص ما لعنوان URL الخاص بك مباشرة في متصفحه. ويحدث هذا عندما يسترد شخص ما عنوان URL الخاص بك من بطاقة أعمال أو كتيب أو إعلان مطبوع أو إعلان راديو وما إلى ذلك، وهذا هو السبب في كونها إستراتيجية جيدة لاستخدام عناوين URL المشفرة.

٤- حملات ليست عبر الإنترنت: إذا كنت تستخدم خيارات الإعلان إلى جانب الحملات القائمة على الويب، فيمكن لبرنامج تحليلات الويب الخاص بك التقاط بيانات الأداء إذا قمت بتضمين آلية لإرسالها إلى موقع الويب الخاص بك. وعادةً ما يكون هذا العنوان URL المخصص الذي قمت بتضمينه في الإعلان الخاص بك (مثل: "www.mycompany.com/offer50")

هــو العنوان الذي يقدم هــؤلاء الزوار لصفحة معينة، ويصبح لديــك الآن بيانات حول عدد المستجيبين لذلك الإعلان عن طريق زيارة موقع الويب الخاص بك.

٥- حملات عبر الإنترنت: إذا كنت تعتمد على اللافتات الإعلانية، أو إعلانات على محرك بحثي، أو حتى حملة بريد إلكتروني، فيمكنك قياس فعالية تلك الحملات الفردية بمجرد استخدام عنوان URL مخصص مشابه لإستراتيجية الحملة التي لا تعتمد على الإنترنت.

الملفات الشخصية للزوار:

وهي إحدى الطرق التي يمكنك من خلالها الاستفادة من تحليلات الويب الخاصة بك وتحويلها إلى أداة تسويقية حقيقية وذلك من خلال التقسيم إلى شرائح. ومن خلال مزج البيانات من تقارير التحليلات المختلفة، ستبدأ في مشاهدة مجموعة متنوعة من الملفات الشخصية للمستخدمين.

١- الكلــمات المفتاحية: فمن خلال تقرير التحليلات الخاص بك، يمكنك معرفة الكلمات المفتاحية التي يستخدمها الزوار في محركات البحث لتحديد موقع الويب الخاص بك. فإذا قمت بتجميع كلماتك الرئيســة بسمات مشــابهة، فإنك ستبدأ في رؤية مجموعات مميزة من الزوار تستخدم موقع الويب الخاص بك. فمثلًا، يمكن أن تشير عبارة البحث المحددة التي تم استخدامها إلى مدى فهمهم لمنتجك أو فوائده. وإذا استخدموا الكلمات التي تعكس وصف المنتج أو الخدمة الخاصة بك، فهذا يعني أنهم قد يكونون بالفعل على علم بالعروض الخاصة بك من الإعلانات الفعالة والكتيبات وما إلى ذلك. وإذا كانت الشروط أكثر عمومية بطبيعتها، فإن زائرك يبحث عن حل لمشكلة ما وقد حصل على موقع الويب الخاص بك. وإذا كانت هذه المجموعة الثانية من الباحثين كبيرة الحجم، فإنك ستحتاج إلى التأكد من أن موقعك يحتوي على مكون تعليمي قوي لإقناعهم بأنهم قد وجدوا إجابتهم ومن ثم نقلهم إلى قناة مبيعاتك.

٢- تجمعات المحتوى: فمن خلال الاعتماد على كيفية تجميع المحتوى الخاص بك، قد تتمكن من تحليل أقسام موقع الويب الخاص بك والتي تتوافق مع منتجات وخدمات وحملات وأساليب تســويقية أخــرى. فإذا كنت تدير الكثير مــن المعارض التجاريــة وزادت الزيارات إلى موقع الويب الخاص بك بحثًا عن المنشــورات الخاصة بمنتج معين، فإن تحليلات الويب الخاصة بك سوف تقوم بتسليط الضوء على النشاط في ذلك القسم.

٣- الجغرافيا: حيث تســمح لــك التحليلات بمعرفة المكان الذي تنبع منــه الزيارات جغرافيًا، بما في ذلك البلد والولاية والمدينة. وقد يكون هذا مفيدًا بالنسبة لك خاصةً إذا كنت تستخدم حملات تستهدف مواقع جغرافية معينة أو كنت تريد قياس مشاهداتك في منطقةٍ ما.

٤- فترات اليوم: فعادةً ما تصل زيارات الويب إلى ذروتها في بداية يوم العمل وخلال وقت الغداء وفي نهاية يوم العمل. ومع ذلك، وليس من المعتاد إيجاد زيارة قوية إلى موقع الويب الخاص بك في وقت متأخر من المساء. ويمكنك تحليل هذه البيانات لتحديد مواعيد تصفح الناس في مقابل الشراء وأيضًا اتخاذ القرارات بشأن الساعات التي يجب أن تقدم فيها خدمة العملاء.

٥- خصائــص صفحــات مطلوبة للبحث: فــإذا قمت بتنظيم حملاتك الإعلانية المختلفة بشــكل صحيــح، فإنــه يمكنك توجيه كل مجموعة مــن مجموعاتك المســتهدفة إلى صفحة مقصودة مختلفة، والتي ستقوم تحليلات الويب الخاصة بك بالتقاطها وقياسها. ومن خلال الجمع بين هذه الأرقام والتوزيعات الســكانية لوسائل الحملة الخاصة بك، يمكنك معرفة النسبة المئوية للزوار التي تناسب كل فئة سكانية.

إحصائيات التحويل:

ســوف تقوم كل منظمة بتعريف «التحويل» وفقًا للأهداف التســويقية الخاصة بها. وتستخدم بعض برامج التحليلات مصطلح هدف (goal) لقياس أهداف موقع ويب معين، سواء كان ذلك عددًا معينًا من الزائرين لصفحة ما أو نموذج تسجيل تام أو عملية شراء عبر الإنترنت.

١- الزوار الجدد: فإذا كنت تعمل على زيادة مســتوى الرؤية، فإنك ستحتاج إلى دراسة المؤشرات في بيانات الزائرين الجدد، حيث يصنف برنامج التحليلات جميع الزوار على أنهم إما جُدد أو مترددين على الموقع.

٢- رجوع الزوار: فإذا كنت تشارك في برامج الولاء أو تقدم منتجًا يحتوي على دورة شراء طويلة، فستساعدك بيانات زوارك العائدين في قياس التقدم في هذا المجال.

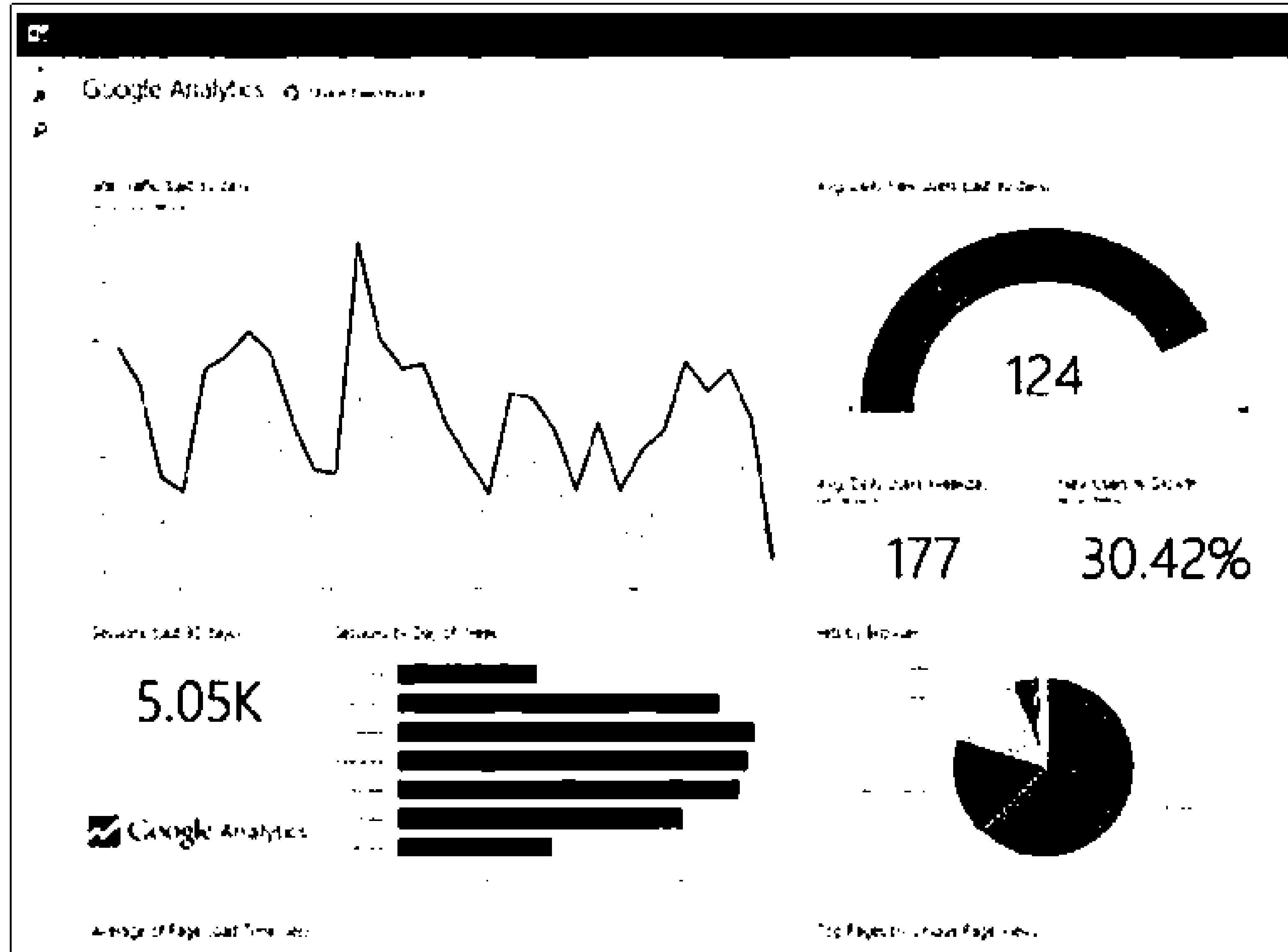
٣- القيادات: فبمجرد تقديم نموذج وتوليد صفحة شكر، فإنك بذلك قد تكون جذبت قائد، حيث تسمح لك تحليلات الويب بحساب معدل الإكمال (أو معدل التخلي) عن طريق قسمة عدد النــماذج المكتملة على عدد زائري الويب الذين جاءوا إلى صفحتك. وتشــير النســبة المئوية للإكمال المنخفض إلى أن الصفحة تحتاج إلى الاهتمام.

٤- المبيعــات / التحويــلات: اعتمادًا على هدف موقع الويب الخاص بــك، يمكنك تحديد "البيع" من خلال صفقة عبر الإنترنت، أو تسجيل مكتمل، أو تقديم عبر الإنترنت، أو أي عدد من أنشطة الويب الأخرى. ومراقبة هذه الأرقام سوف تنبهك إلى أي تغييرات (أو نجاحات!) تحدث في مراحل أعلى.

٥- معــدلات المغادرة / الخروج: وهي لا تقل أهمية عن معدلات التحرك من موقع الويب الخاص بــك، ويقصد بها معدلات الأشــخاص الذين بدأو عملية التحــرك في الموقع ثم قاموا بإنهائها، أو

وصلوا إلى موقع الويب الخاص بك وتركوه بعد صفحة أو اثنتين. ففي الحالة الأولى، ستحتاج إلى تحليل المكان الذي أنهى فيه الزائر العملية وما إذا كان هناك عدد من الزائرين يغادرون في المكان نفسه، ومن ثَم التحقيق في الوضع من أجل اتخاذ القرار. وفي الحالة الأخيرة، يشير معدل الخروج المرتفع من موقع ويب أو صفحة ما بشكل عام إلى وجود مشكلة في التوقعات. ويقوم الزائرون بالنقر على موقع الويب الخاص بك استنادًا إلى بعض الرسائل الواردة في الإعلان، والعرض التقديمي، وما إلى ذلك، ومن المتوقع استمرار هذه الرسالة بعض الوقت. وينبغي عليك أن تتأكد من أنك تعلن الرسالة التي يستطيع موقعك على الويب أن يعززها ويسلمها.

ويتضمن كل بند من هذه العناصر مقاييس يمكن إنشاؤها لمؤسستك. كما يمكنك إنشاء لوحة معلومات أسبوعية تتضمن أرقامًا أو نسبًا محددة تشير إلى المكان الذي تنجح فيه - أو تسلط الضوء على التحدي التسويقي الواجب معالجته. وعندما يتم تقييم هذه المقاييس باستمرار واستخدامها مرتبطةً ببيانات تسويقية أخرى متاحة، فمن الممكن أن يقودك ذلك إلى برنامج تسويق كمي للغاية. ويعرض الشكل (٥-١٤) لوحة تحكم في تحليلات الويب تم إنشاؤها باستخدام أدوات Google Analytics المتاحة مجانًا.



شكل ٥-١٤: عينة لوحة معلومات لتحليلات الويب

أسئلة مراجعة على القسم ٥-٩:

١- ما هي الأنواع الثلاثة للبيانات التي يتم توليدها من خلال زيارات صفحة الويب؟

٢- ما هو تحليل مسار النقر؟ وفيم يستخدم؟

٣- ما هي التطبيقات الرئيسة للتنقيب في الويب؟

٤- ما هي مقاييس تحليلات الويب شائعة الاستخدام؟ وما أهمية المقاييس؟

٥-١٠ التحليلات الاجتماعية:

يتفاوت معنى التحليلات الاجتماعية من شخص لآخر، وفقاً لنظرته إلى العالم، وإلى مجال دراسته. فعلى سبيل المثال، يشير تعريف القاموس للتحليلات الاجتماعية إلى وجهة نظر فلسفية وضعها المؤرخ والفيلسوف الدنماركي Lars-Henrik Schmidt في ثمانينيات القرن العشرين. ويعد الهدف النظري من وجهة النظر هذه هو socius، وهو نوع من "الشمولية" التي لا تمثل حسابًا عالميًّا ولا مشاركة مجتمعية بين كل عضو في كيان ما (Schmidt، 1996). وهكذا، تختلف التحليلات الاجتماعية عن كل من الفلسفة التقليدية وعلم الاجتماع، والتي قد ينظر إليها على أنها المحاولات الرامية إلى توضيح الخلاف بين الفلسفة وعلم الاجتماع.

ويختلف تعريفنا للتحليلات الاجتماعية إلى حد ما. فبدلًا من التركيز على الجزء «الاجتماعي» (كما هو الحال في التعريف الفلسفي له)، فنحن نهتم أكثر بجزء «التحليلات» في المصطلح. وقد قامت شركة Gartner (وهي شركة استشارية عالمية معروفة في مجال تقنية المعلومات) بتعريف التحليلات الاجتماعية على أنها «رصد وتحليل وضبط وتفسير التفاعلات الرقمية والعلاقات بين الأشخاص والمواضيع والأفكار والمحتوى» (gartner.com/it-glossary/social-analytics/). وتشمل التحليلات الاجتماعية التنقيب في محتوى النص الذي تم إنشاؤه في وسائل التواصل الاجتماعي (مثل تحليل المشاعر، ومعالجة اللغات الطبيعية NLP) وتحليل الشبكات الاجتماعية (مثل تحديد المؤثر، والتنميط، والتنبؤ) وذلك بغرض الحصول على إحصاءات حول السلوكيات الحالية والمستقبلية للعملاء الحاليين والمحتملين، وحول شعورهم بالإعجاب أو عدم الإعجاب تجاه منتجات وخدمات الشركة. واستنادًا إلى هذا التعريف والممارسات الحالية، يمكن تصنيف التحليلات الاجتماعية إلى فرعين مختلفين - غير أنه ليس بالضرورة أن يكون هذين الفرعين حصريين - وهما: تحليل الشبكات الاجتماعية (SNA) وتحليلات وسائل التواصل الاجتماعي.

تحليل الشبكة الاجتماعية:

تعتبر الشــبكة الاجتماعية هيــكل اجتماعي يتكون من أفراد / أشــخاص (أو مجموعات من الأفــراد أو المنظمات) ترتبــط ببعضها البعض بنوع من الروابــط / العلاقات؛ حيث يقدم منظور الشــبكة الاجتماعية طريقة شــاملة لتحليل بنية وديناميكيات الكيانات الاجتماعية. وتَســتخدم دراسة هذه الهياكل نظام الحسابات القومية لتحديد الأنماط المحلية والعالمية، وتحديد الكيانات المؤثرة، ودراســة ديناميكية الشــبكة. والجدير بالذكر أن الشــبكات الاجتماعية وتحليلها هي في الأســاس مجال متعدد التخصصات نشأ من علم النفس الاجتماعي، وعلم الاجتماع، والإحصاءات، ونظرية الرســم البياني. ويعود تاريخ تطوير وإضفاء الطابع الرســمي على المدى الرياضي لنظام الحســابات القومية إلى الخمســينيات، في حين أن تاريخ تطوير النظريات والأســاليب الأساســية للشــبكات الاجتماعية يعود إلى ثمانينيات القرن العشرين(Scott & Davis, 2003). ويُعد نظام الحســابات القومية الآن واحداً من النماذج الرئيســة في تحليلات الأعمال، وذكاء المستهلك، وعلم الاجتماع المعاصر، كما يستخدم أيضًا في عدد من العلوم الاجتماعية والرسمية الأخرى.

إن الشــبكة الاجتماعيــة هي بنيان نظــري مفيد في العلوم الاجتماعية لدراســة العلاقات بين الأفــراد والجماعات والمنظمات، أو حتى مجتمعات بأكملها (الوحدات الاجتماعية). ويســتخدم المصطلــح لوصف البنية الاجتماعية التــي تحددها مثل هذه التفاعلات، حيث تمثل الروابط التي ترتبــط بها أي وحدة اجتماعية معينة تقارب مختلف جهــات التواصل الاجتماعي لتلك الوحدة. وبشكل عام، فإن الشبكات الاجتماعية ذاتية التنظيم، ناشئة ومعقدة، بحيث يظهر نمط متماسك عالميًا من التفاعل المحلي للعناصر (الأفراد ومجموعات الأفراد) التي تشكل النظام.

وفيما يلي بعض أنواع الشبكات الاجتماعية النموذجية المرتبطة بأنشطة الأعمال:

- **شــبكات الاتصالات:** تُعتبر دراســات الاتصــال في الغالب جزءًا من كلٍّ مــن العلوم الاجتماعية والإنســانية، وتعتمد إلى حد كبير على مجالات مثل علم الاجتماع وعلم النفس وعلم الإنســان (الأنثروبولوجــي) وعلم المعلومات والأحياء (البيولوجي) والعلوم السياســية والاقتصاد. وتفيد العديد من مفاهيم الاتصالات في وصف عملية نقل المعلومات من مصدر إلى آخر وبالتالي يمكن تمثيلها كشــبكة اجتماعية، حيث تستفيد شركات الاتصالات من هذا المصدر الغني بالمعلومات لتحسين ممارسات الأعمال الخاصة بها وكذلك تحسين علاقات العملاء.
- **شبكات المجتمع:** يشير مصطلح المجتمع تقليديًا إلى موقع جغرافي معين، وكانت دراسات العلاقات المجتمعية تتعلق بمن تحدثوا، وترابطوا، وتاجروا، وحضروا نشاطات اجتماعية مع بعضهم البعض. ومــع ذلك، هناك اليوم مجتمعات "إلكترونية" موســعة تم تطويرها مــن خلال أدوات التواصل

الاجتماعي وأجهزة الاتصالات. مثل هذه الأدوات والأجهزة تنتج باستمرار كميات كبيرة من البيانات، والتي يمكن استخدامها من قبل الشركات لاكتشاف معلومات نفيسة وقابلة للتطبيق.

- **الشبكات الإجرامية:** في علم الجريمة وعلم الاجتماع المدني، تم توجيه اهتمام كبير للشبكات الاجتماعية من بين الممثلين الإجراميين. فمثلًا، قد تؤدي دراسة جرائم القتل التي تقوم بها العصابات والأنشطة غير القانونية الأخرى - باعتبارها سلسلة من عمليات التناوب بين العصابات - إلى فهم أفضل لهذه الأنشطة الإجرامية ومن ثم العمل على منعها. والآن وبعد أن أصبحنا نعيش في عالم مترابط إلى حد كبير (بفضل الإنترنت)، تقوم وكالات الأمن بمراقبة / متابعة الكثير من تشكيلات الشبكات الإجرامية وأنشطتها باستخدام أدوات وتكتيكات الإنترنت الحديثة. وعلى الرغم من أن الإنترنت قد غيّر شكل شبكات الإجرام والوكالات المنوطة بتطبيق القانون، إلا أن النظريات الاجتماعية والفلسفية التقليدية لا تزال تطبق إلى حد كبير.
- **شبكات الابتكار:** تركز دراسات الأعمال على نشر الأفكار والابتكارات في بيئة الشبكات على انتشار الأفكار واستخدامها بين أعضاء الشبكة الاجتماعية. وتكمن الفكرة في فهم سبب كون بعض الشبكات أكثر إبداعًا، ولماذا تتبنّى بعض المجتمعات تلك الأفكار والابتكارات في وقت مبكر (مثل دراسة أثر هيكل الشبكة الاجتماعية في التأثير على انتشار الابتكار والسلوك الإبداعي).

مقاييس تحليل الشبكة الاجتماعية:

إن نظام تحليل الشبكة الاجتماعية SNA هو الفحص المنتظم للشبكات الاجتماعية. وينظر نظام تحليل الشبكة الاجتماعية إلى العلاقات الاجتماعية من حيث نظرية الشبكة، والتي تتكون من عُقد nodes (تمثل الأفراد أو المنظمات داخل الشبكة) والعلاقات / الاتصالات (التي تمثل العلاقات بين الأفراد أو المنظمات، مثل الصداقة أو القرابة أو الوضع التنظيمي). وغالبًا ما يتم تمثيل هذه الشبكات باستخدام الرسوم البيانية للشبكة الاجتماعية، حيث يتم تمثيل العُقد (نقاط الالتقاء) كنقاط ويتم تمثيل الروابط كخطوط. وتقدم الحالة العملية (٥-٨) مثالًا مثيرًا للتحليلات الاجتماعية متعددة القنوات.

حالة عملية ٥-٨

فودكا تيتو تؤسس الولاء للعلامة التجارية بإستراتيجية اجتماعية حقيقية

إذا كان على الفودكا المصنوعة يدويًّا من تيتو Tito تحديد مقياس واحد لوسائل التواصل الاجتماعي يعكس مهمتها بدقة أكبر، فربما يكون المشاركة. إن التواصل مع محبي الفودكا بطريقة شاملة وأصيلة هو أمر تأخذه تيتو Tito بجدية بالغة، وتعكس الإستراتيجية الاجتماعية للعلامة التجارية تلك الرؤية.

وقد كان تأسيسـها منذ ما يقرب من عقدين من الزمان، وقد ساعد ظهور وسائل التواصل الاجتماعي في تعزيز العلامة التجارية من خلال لعب دور أساسي في إشراك الجمهور وزيادة الوعي بالعلامة التجارية. وفي مقابلة مع صاحب المشروع، ذكر أن Bert Beveridge مؤسس شركة تيتو «Tito"، قد نسب الفضل لوسائل التواصل الاجتماعي لتمكينه من التنافس على مساحة العرض على الرفوف مع علامات تجارية أخرى للخمور أكثر شهرة، حيث قال Beveridge لصاحب المشروع: "وسائل التواصل الاجتماعي هي منصة كبيرة للترويج لعلامة تجارية من خلال الكلمة الشفوية، لأنها ليست فقط لمن لديه أكبر مكبر للصوت".

ومع نضج شركة تيتو Tito ظل الفريق الاجتماعي مخلصًا للقيم التأسيسية للعلامة التجارية ويستخدم كلًا من Twitter وInstagram بنشاط لإجراء محادثات فردية وتواصل مع المتحمسين للعلامة التجارية. وقالت Katy Gelhausen، منسقة شبكة الإنترنت ووسائل التواصل الاجتماعي: "نحن لم ننظر أبدًا لوسائل التواصل الاجتماعي على اعتبار أنها وسيلة أخرى للإعلان". وتابعت: "نحن موجودون على هذه الوسائل الاجتماعية حتى يتمكن عملائنا من التحدث إلينا".

وللوصول لهذه النهاية تستخدم تيتو Tito وسائل التواصل الاجتماعي الحديثة لفهم جو الصناعة، وتطوير علامة تجارية اجتماعية متسقة، وإنشاء حوار مع جمهورها. ونتيجة لذلك، فقد قامت تيتو Tito بتنمية مجتمعاتها عضويًا على كل من Twitter بنسبة ٤٣,٥٪ وInstagram بنسبة ١٢,٦٪، وذلك في غضون ٤ أشهر.



الإخبار بإستراتيجية التسويق الموسمية المتكاملة:

يمثل برنامج كوكتيل الربع سنوي لتيتو Tito جزءًا رئيسًا من إستراتيجية التسويق المتكاملة للعلامة التجارية؛ حيث يتم تطوير وصفة كوكتيل في كل ربع سنة وتوزيعها من خلال مبادرات التسويق عبر الإنترنت في تيتو Tito.

ومن المهم بالنسبة لتيتو Tito ضمان توافق الوصفة مع تركيز العلامة التجارية فضلًا عن الاتجاه إلى صناعة أكبر. ولذلك، تستخدم Gelhausen الكلمات المفتاحية لعلامة Sprout لمراقبة اتجاهات الصناعة وملفات تعريف نكهة الكوكتيل. تقول Gelhausen: "لقد كان جهاز Sprout أداة مهمة جدًا للرصد الاجتماعي. ويُعد Inbox طريقة جيدة للاحتفاظ بأعلى علامات التصنيف والاطلاع على المؤشرات العامة في جدول واحد".

ويتم تقديم هذه الدروس إلى فريق عمل المزيج الداخلي في تيتو Tito واستخدامها لضمان توصيل نفس الوصفة الربع سنوية إلى فريق مبيعات العلامة التجارية وعبر قنوات التسويق. تقول Gelhausen: "سواء كنت تشرب تيتو Tito في البار أو تشتريه من متجر لبيع الخمور أو تتابعنا على مواقع التواصل الاجتماعي فإنك تحصل على نفس الكوكتيل الربع سنوي".

يضمن البرنامج حصول الشخص على تجربة متسقة للعلامة التجارية، في كل نقطة اتصال للمستهلك، وهذا الاتساق أمر حيوي. وفي الواقع، فإنه وفقًا لدراسة أجرتها شركة Infosys حول تجربة التسوق التي توفرها omnichannel، فإن ٣٤٪ من المستهلكين يرون أن الاتساق بين القنوات هو سبب إنفاقهم أكثر مع علامة تجارية. وعلى الجانب الآخر ذكر ٣٩٪ أن التناقض يُعد سببًا كافيًا لإنفاق مبالغ أقل.

ويبدأ جمع رؤى الصناعة في شركة تيتو Tito بمراقبة اجتماعية على Twitter وInstagram من خلال Sprout. غير أن الإستراتيجية الاجتماعية للعلامة التجارية لا تتوقف عند هذا الحد، حيث تستخدم تيتو Tito النظام الأساسي بشكل يومي للتواصل فعليًا مع عملائها، مع الحفاظ على جذورها.

ويقوم صندوق البريد الذكي في Sprout بعرض حسابات Twitter وInstagram الخاصة بتيتو Tito في خلاصة واحدة ومتماسكة، مما يساعد Gelhausen على إدارة الرسائل الواردة والتحديد بشكل سريع أي منها يتطلب استجابة.

يقول Gelhausen: "يتيح لنا برنامج Sprout البقاء على اتصال دائم بالمحادثات التي نجريها مع متابعينا. وأنا أحب إمكانية التفاعل بسهولة مع المحتوى من حسابات متعددة في مكان واحد".

انتشار الكلمة على تويتر:

وتُعد الطريقة التي تتبعها تيتو Tito على تويتر بسيطة، حيث إنها تتمثل في الانخراط في محادثات شخصية فردية مع المعجبين. ويعد الحوار قوة دافعة للعلامة

التجارية، وعلى مدار ٤ أشـهر، كانت نسـبة ٨٨٪ من التغريدات التي تم إرسـالها تمثل ردودًا على الرسائل الواردة.

وقد أدى استخدام Twitter كخط اتصال بين تيتو Tito ومعجبيها إلى زيادة التفاعل بنسـبة ١٦٢,٢٪، بالإضافة إلى زيادة المتابعين بنسـبة ٤٣,٥٪. وما يثير الإعجاب أكثر هو إنهاء تيتو Tito هذا الربع السـنوي بواقع ٥٣٨,٣٠٦ انطباعًا من الأعضاء - وهو ارتفاع بنسـبة ٨١٪. كما يتم تطبيق إستراتيجية مشـابهة على Instagram، حيث تستخدمها تيتـو Tito لتعزيز وتقويـة العلاقة مع المعجبين وذلك من خـلال نشر الصور ومقاطع الفيديو لأفكار الوصفات الجديدة وأحداث العلامة التجارية والمبادرات.

الاستيلاء على الحفل على Instagram:

تقوم تيتو Tito في الأسـاس بنـشر محتوى نمط الحياة على Instagram وتشـجع المتابعين لدمج العلامة التجارية في المناسـبات اليومية. كما تسـتخدم تيتو Tito أيضًا النظام الأسـاسي للترويج لجهودها التسـويقية ولإخبار قصة علامتهـا التجارية. ويعثر الفريـق على قيمـة في تقريـر Instagram الخاص بـ Sprout، والتي تسـاعدهم على تحديد أكثر الأشـياء التي تحظى بالمشـاركة على وسـائل التواصل وتلقي أكبر قدر من التفاعل وتحليل تركيبة الجمهور ومدى زيادتهم، والغوص في أنماط النشر بشكل أعمق، وتقييم أداء الوسـم (hashtag) الصادر. وفي هذا الصدد يقول Gelhausen: "في ضوء الخلاصة المخصصة الجديدة في Instagram، من المهم أن ننتبه لما يحدث بالفعل".

وقـد تمكنت تيتو Tito باسـتخدام تقريـر Instagram من قياس أثر إسـتراتيجيتها التسـويقية Instagram وتعديـل مقاربتها وفقًا لذلك. ومن خلال اسـتخدام الشـبكة كطريقة أخرى للتفاعل مع المعجبين، نمت علامتها التجارية بشكل مطرد مع نمو جمهورها. وفي غضون ٤ أشـهر شـهدت TitosVodka@ زيادة في المتابعين بنسبة ١٢,٦٪، وزيادة في التفاعل بنسـبة ٣٧,١٪. كما حصل كل محتوى منشـور على متوسط ٥٣٤ تفاعلًا، كما نما الوسم (hashtag) الخاص بالعلامة التجارية titoshandmadevodka# بنسبة ٣٣٪.

من هنا يؤدي إلى أين؟

تعد وسـائل التواصل الاجتماعي استثمارًا متواصلًا في الوقت والاهتمام. وستواصل تيتـو Tito النشـاط الذي شـهدته علامتها التجارية من خلال تقسـيم كل ربع سـنة في حملتهـا الخاصـة. يقـول Gelhausen: "نحـن دائمًـا نصبح أكـثر ذكاءً من خلال

إستراتيجياتنا الاجتماعية ونحرص على أن يكون ما ننشره ملائمًا ونتلقى صداه". وسيظل استخدام التواصل الاجتماعي مع المعجبين بطريقة ثابتة وأصلية ولا تنسى حجر الزاوية في جهود التسويق الرقمي للعلامة التجارية. وستواصل تيتو Tito تعزيز مجتمع محبيها loyalists، باستخدام مجموعة أدوات إدارة وسائل التواصل الاجتماعي في Sprout.



ملاحظات:

- توجد زيادة بنسبة ١٦٢٪ في التفاعل الأساسي على Twitter.

- وزيادة بنسبة ٨١٪ في الانطباعات الأساسية على Twitter.

- وزيادة بنسبة ٣٧٪ في التفاعل على Instagram.

أسئلة للمناقشة:

١- كيف يمكن استخدام تحليلات وسائل التواصل الاجتماعي في صناعة المنتجات الاستهلاكية؟

٢- ما هي في رأيك التحديات الرئيسة والحلول والنتائج المحتملة لتطبيق تحليلات وسائل التواصل الاجتماعي على المنتجات الاستهلاكية وشركات الخدمات؟

Source: SproutSocial Case Study, "Tito's Vodka Establishes Brand insights/case-studies/titos/(accessed July 2016).

وبمرور الأعوام، تم تطوير مقاييس (أو قياسات) مختلفة لتحليل هياكل الشبكات الاجتماعية من وجهات نظر مختلفة. وغالبًا ما يتم تجميع هذه المقاييس ضمن ثلاث فئات: الروابط والتوزيعات والتقسيم إلى قطاعات.

الروابط:

حب الأشـياء المتشـابهة: وهو مدى قيام الجهات الفاعلة بتشـكيل روابط مع آخرين بينهم تشـابه مع غيرهم ممن يختلفون عنهم. ويمكن تعريف التشـابه حسـب الجنس أو النسب أو العمر أو المهنة أو التحصيل العلمي أو الحالة أو القيم أو أي سمة بارزة أخرى.

التعدد: وهو عدد من نماذج المحتوى المضمنة في رابط، مثل: شخصين من الناس، وهما صديقان، بالإضافة إلى أنهما يعملان معًا مجموعة متعددة من اثنين. يرتبط التعدد بقوة العلاقة.

تبادل العواطف / المعاملة بالمثل: هو مدى قيام اثنين من الممثلين بتبادل علاقات الصداقة أو غيرها من التفاعلات بين كل منهما.

إغلاق الشـبكة: وهو مقياس لإكمال العلاقات الثلاثية، حيث إن افتراض الفرد لإغلاق الشبكة (أي أن أصدقاء أصدقائه هم أصدقاؤه أيضًا) يسـمى التعدي Transitivity. والتعدي هو نتيجة للخاصية الفردية أو الظرفية للحاجة إلى الإغلاق الإدراكي.

التجاور: وهو الميل للحصول على مزيد من الروابط مع الآخرين ذوي الموقع الجغرافي المقرب.

التوزيعات:

الجسر: وهو الشـخص الذي يملأ فجوة هيكلية عن طريق روابطه الضعيفة، وهو الذي يوفر الرابط الوحيد بين فردين أو مجموعة. كما يتضمن أيضًا أقصر الطرق عندما يكون المدى الأطول غير ممكن بسبب وجود خطر كبير لتشويه الرسائل أو فشل التسليم.

المركزية: وتشير إلى مجموعة من المقاييس التي تهدف إلى تحديد أهمية أو تأثير (في مجموعة متنوعـة من الحواس) عقـدة معينة (أو مجموعة) داخل الشـبكة. ويوجد عـدة طرق لقياس المركزيـة من أشـهرها المركزية البينية، ومركزية التقارب، ومركزية القيمـة الذاتية، ومركزية ألفا، بالإضافة إلى مركزية الدرجة.

الكثافة: وهي نسبة الروابط المباشرة في الشبكة بالنسبة إلى العدد الإجمالي الممكن.

المسافة: وهي الحد الأدنى لعدد الروابط المطلوبة لربط جهتين محددتين.

الثقـوب الهيكلية: وهـي غياب الروابط بين جزئين من الشـبكة، حيث إن العثور على ثقب هيكلي واسـتغلاله من الممكن أن يؤدي إلى منح صاحب المشروع ميزة تنافسـية. وقد تم تطوير هذا المفهوم من قبل عالم الاجتماع Ronald Burt ويشـار إليه أحيانًا على أنه مفهوم بديل لرأس المال الاجتماعي.

قـوة الربط: وتعرف بأنها مزيج خطي من الوقت، والكثافة العاطفية، والحميمية، والمعاملة بالمثـل (أي التبادليـة). وتترابـط العلاقات القوية بكل من حب الأشـياء المتشـابهة، والتجاور، والتعدي، في حين تترابط العلاقات الضعيفة مع الجسور.

التقسيم إلى قطاعات:

الزُمَـر والدوائر الاجتماعية: ويتم تعريف المجموعات عـلى أنها زمر إذا كان كل فرد مرتبطًا بشـكل مباشر بجميع الدوائر الفردية أو الاجتماعية الأخرى مع مسـتوى صرامة أقل من الاتصال المباشر، وهذا تعريف غير دقيق، أو تعرف على أنها مجموعات متماسـكة هيكليًا في حالة الرغبة في مستوى دقة أكبر.

معامل التجميع (Clustering Coefficient): وهو مقياس لاحتمال كون عضوين من العقدة مرتبطين. ويشير معامل التجميع الأعلى إلى درجة أكبر من الطائفية.

التماسـك: وهو مـدى ارتباط الممثلين مباشرة ببعضهم البعض عن طريق روابط متماسـكة. ويشير التماسك الهيكلي إلى الحد الأدنى لعدد الأعضاء الذين إذا تمت إزالتهم من مجموعة ما فإن ذلك من شأنه أن يؤدي إلى فك المجموعة.

تحليلات وسائل التواصل الاجتماعي:

تشـير وسـائل التواصل الاجتماعي إلى التقنيات المواتية للتفاعلات الاجتماعية بين الأشـخاص الذين يقومون بإنشـاء ومشـاركة وتبـادل المعلومات والأفكار والآراء في المجتمعات والشـبكات الافتراضيـة، وهـي عبارة عن مجموعة من تطبيقات البرامج المسـتندة إلى الإنترنت والتي تعتمد على الأسـس الأيديولوجية والتقنية للويب ٢,٠ والتي تسمح بإنشاء وتبادل المحتوى الذي ينشئه المسـتخدم (Kaplan & Haenlein, 2010). وتعتمد وسـائل التواصـل الاجتماعي على تقنيات المحمـول وغيرهـا من التقنيات القائمة على الويب لإنشـاء منصات تفاعلية بشـكل قوي للأفراد والمجتمعات لمشـاركة المحتوى الذي ينشئه المسـتخدم وإنشاء المحتوى ومناقشته وتعديله. كما أنها تقدم تغييرات كبيرة في التواصل بين المنظمات والمجتمعات والأفراد.

ومنذ ظهورها في أوائل التسـعينيات شهدت تقنيات الشبكات الاجتماعية القائمة على الويب تحسـنًا كبيرًا في كل من الجـودة والكمية. وتتخذ هذه التقنيات العديد من الأشـكال المختلفة، بما في ذلك المجـلات عبر الإنترنت، ومنتديات الإنترنت، وسـجلات الويب، والمدونات الاجتماعية، والمدونـات الصغيرة، والويكي، والشـبكات الاجتماعية، والحلقات الصوتيـة podcasts، والصور، والفيديو، وتقييمات / تصنيفات المنتجات / الخدمات. ومن خلال تطبيق مجموعة من النظريات

في مجال أبحاث وسائل التواصل (الحضور الاجتماعي، ثراء وسائل التواصل) والعمليات الاجتماعية (التقديـم الذاتي، والإفصاح الذاتي)، أنشـأ كل مـن Kaplan وHaenlein (2010) نظام تصنيف لوسائل التواصل الاجتماعي بست أنواع مختلفة منها : (١) المشاريع التعاونية (مثل ويكيبيديا)، (٢) مدونـات والمدونـات الصغيرة (مثل تويـتر)، (٣) مجتمعات المحتوى (مثـل: يوتيوب)، (٤) مواقع الشـبكات الاجتماعية (مثل Facebook)، (٥) عـالم الألعاب الافتراضية (مثل: World of Warcraft)، (٦) عوالم اجتماعية افتراضية (مثل: Second Life).

وتختلف وسائل التواصل الاجتماعي القائمة على الويب عن وسائل التواصل التقليدية / الصناعية، مثل الجرائد، والتلفاز، والأفلام، حيث إنها رخيصة نسـبيًّا ويمكن الوصول إليها لتمكين أي شـخص (حتى الأفراد العاديين) من نشر أو الوصول إلى / اسـتهلاك المعلومات. وتتطلب وسـائل التواصل الصناعيـة عمومًـا موارد مهمـة لنشر المعلومات، حيـث إن المقالات (أو الكتـب) تمر في معظم الحالات بالعديد من التنقيحات قبل نشرها. وفيما يلي بعض الخصائص الأكثر انتشارًا التي تساعد على التمييز بين وسائل التواصل الاجتماعي والصناعي (Morgan, Jones, & Hodges, 2010):

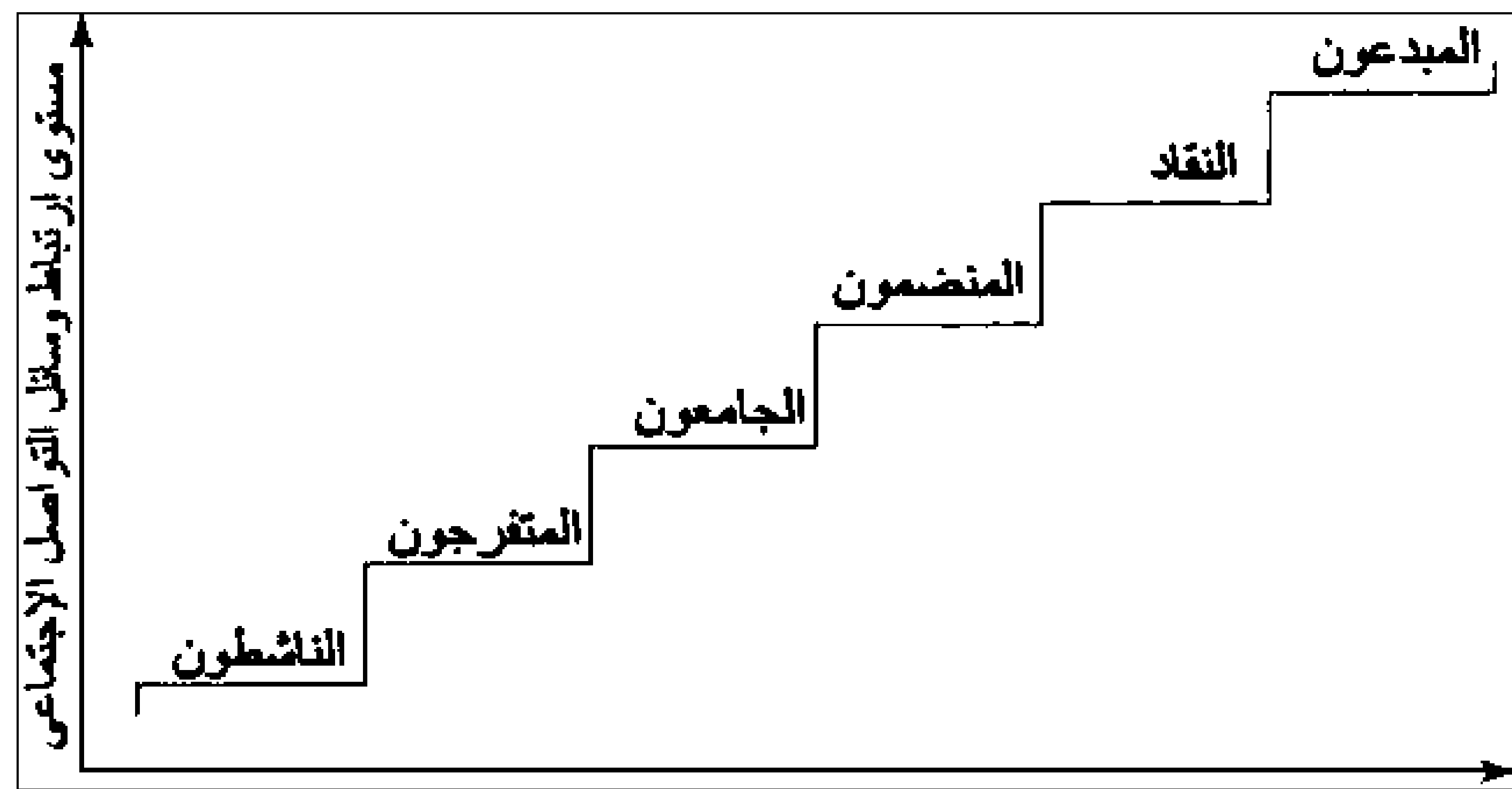
- الجـودة: ففي مجال النشر الصناعي - بواسـطة ناشر - فإن النطـاق النموذجي للجودة يكون أضيق بشكل جوهري من الأسواق المتخصصة غير المغطاة. ويتمثل التحدي الرئيس الذي يمثله المحتوى في مواقع الشـبكات الاجتماعية في حقيقة أن توزيع الجودة له تباين كبير: من عناصر عالية الجودة إلى محتوى منخفض الجودة، وأحيانًا مُسيء.
- الوصول: توفر كل من تقنيات وسـائل التواصل الصناعي والاجتماعي نطاقًا واسـعًا وقادرًا على الوصول إلى جمهور عالمي. ومع ذلك، فعادةً ما تستخدم وسائل التواصل الصناعي إطارًا مركزيًّا للتنظيم والإنتاج والنشر، في حين أن وسـائل التواصل الجتماعـي بطبيعتها أكثر لامركزية وأقل تسلسلًا وتميزًا بنقاط متعددة للإنتاج والمنفعة.
- التكرار: حيث إن تحديث وإعادة النشر على منصات وسـائل التواصل الاجتماعي يُعد أسـهل وأسرع وأرخص، وذلك بالمقارنة مع وسـائل التواصل الصناعية وبالتالي فإنه يمارس بشـكل أكثر تكرارًا، مما يؤدي إلى الحصول على محتوى أعذب.
- إمكانية الوصول: عادةً ما تكون وسائل الإنتاج الخاصة بوسائل التواصل الصناعية حكومية و / أو شركات (مملوكة ملكية خاصة) ومكلفة، في حين أن أدوات وسائل التواصل الاجتماعي متاحة للجمهور بشكل عام بتكلفة قليلة أو بدون تكلفة.
- سـهولة الاستخدام: يتطلب إنتاج وسـائل التواصل الصناعية عادةً مهارات متخصصة وتدريب. وعـلى العكس من ذلك، فإن معظم إنتاج وسـائل التواصل الاجتماعي لا يتطلب سـوى إعادة

تفســير متواضع للمهارات القائمة؛ ومن الناحية النظرية، فإن أي شخص لديه إمكانية، يستطيع تشغيل وسائل إنتاج وسائل التواصل الاجتماعي.

- فوريــة: قد يكــون الفارق الزمني بين الاتصالات التي تنتجها وســائل التواصل الصناعية طويلًا (أســابيع أو شهور أو حتى سنوات) مقارنةً بوســائل التواصل الاجتماعي (التي يمكن أن تكون قادرة على الاستجابة بصورة فورية تقريبًا).
- قابلية التحديث: لا يمكن تغيير وســائل التواصل الصناعية بمجرد إنشــائها (بمجرد طباعة مقالة المجلة وتوزيعها، ولا يمكن إجراء تغييرات على نفس المقالة)، في حين يمكن تغيير وسائل التواصل الاجتماعي على الفور تقريبًا من خلال التعليقات أو التحرير.

كيف يستخدم الناس وسائل التواصل الاجتماعي؟

لا يقتصر نمو مواقع الشــبكات الاجتماعية على عددها فحســب، بل تنمو أيضًا درجة ارتباطها بالقناة. وقد قدم كل من Brogan وBastone (2011) نتائج بحث تقسم المستخدمين حسب مدى نشاطهم في استخدام وســائل التواصل الاجتماعي والتطور المتتبع لهذه القطاعات من المستخدمين بمرور الوقت. وقد أدرجوا ستة مستويات مشاركة مختلفة (الشكل ٥-١٥).



شكل ٥-١٥: تطور مشاركة مستخدم وسائل التواصل الاجتماعي

ووفقًا لنتائج البحث، فقد كان مجتمع المستخدمين عبر الإنترنت يتحول بشكل متزايد باتجاه المشــاركة هرمية الشكل. والتغير الملحوظ بشدة هو الحادث فيما بين غير النشطين، حيث تضم هذه الفئة ٤٤٪ من مستخدمي الإنترنت. وبعد ذلك بعامين، قفز أكثر من نصف هؤلاء الخاملين إلى وســائل التواصل الاجتماعي بشــكل أو بآخر. وقد أشــار Bastone إلى أن: "الآن حوالي ٨٢٪ من مســتخدمي الإنترنــت البالغين ضمن واحدة من الفئات العليا، لقد وصلت وســائل التواصل الاجتماعي إلى حالة من التبنِّي الجماعي".

وقال Bastone: "تشـير وسـائل التواصل الاجتماعي إلى الطرق المنهجية والعلمية لاستهلاك الكمية الهائلة من المحتوى الذي أنشأته وسائل التواصل الاجتماعي على شبكة الإنترنت، والأدوات والتقنيات من أجل تحسـين قدرة المؤسسة على المنافسة. وقد أصبحت تحليلات وسائل التواصل الاجتماعـي قـوة جديدة في المنظمات حول العالم بشـكل سريع، مما يسـمح لهـا بالوصول إلى المسـتهلكين وفهمهم كما لم يحدث من قبل. وبالإضافة إلى ذلك، فقد أصبحت وسـائل التواصل الاجتماعي في العديد من الشركات، أداة للإستراتيجيات المتكاملة للتسويق والاتصالات.

يوفـر النمو الهائـل لوسـائل التواصل الاجتماعـي، مـن المدونـات، وFacebook، وتويتر، إلى LinkedIn، ويوتيوب، وأدوات التحليلات التي تسـتفيد من هـذه المصادر الغنية بالبيانات، للمنظـمات فرصة الدخول في محادثات مع ملايين العملاء حول العالم كل يوم. وهذه الميزة هي السبب في أن ما يقرب من ثلثي الشركات الـ ١٢٠٠ التي شاركت في دراسة استقصائية حديثة - أجرتها هارفارد بيزنس ريفيو أناليتيكس سيرفيسز (Harvard Business Review Analytic Services) - قالوا إنهم إما يسـتخدمون حاليًا قنوات وسـائل التواصل الاجتماعي أو لديهم خطط وسـائل التواصـل الاجتماعي في الأعـمال (Harvard Business Review, 2010)، ولكن ما زال الكثيرون يقولون إن وسـائل التواصل الاجتماعي هي تجربة، حيث يحاولون فهم كيفية الاسـتخدام الأمثل للقنوات المختلفة، وقياس فعاليتها، ودمج وسائل التواصل الاجتماعي في إستراتيجيتهم.

قياس تأثير وسائل التواصل الاجتماعي:

بالنسـبة للمؤسسـات - سـواء كانت صغيرة أو كبيرة - فهناك إحصاءات قيّمة مخفية داخل المحتوى الذي أنشـأه المسـتخدم على مواقع التواصل الاجتماعي، ولكـن كيف يمكنك أن تبرزها مـن بين عشرات مواقـع المراجعة وآلاف المدونـات والملايين من مشـاركات Facebook وملايين التغريدات؟ وإذا أمكنك القيام بذلك، فكيف تقيس تأثير جهودك؟ وهذه الأسئلة يمكن معالجتها مـن خلال تمديد التحليلات لتقنيات وسـائل التواصل الاجتماعـي. وبمجرد أن تحدد هدفك من وسـائل التواصل الاجتماعي (مـا تريد تحقيقه)، فإن هناك العديد من الأدوات لمسـاعدتك على الوصول إلى هذا الهدف، وهذه الأدوات التحليلية تقع عادةً في ثلاث فئات واسعة:

- التحليلات الوصفية: تسـتخدم إحصائيات بسـيطة لتحديد خصائص واتجاهات النشـاط، مثل عـدد المتابعين لديك، وعدد المراجعات التي تم إنشـاؤها على Facebook، والقنوات التي يتم استخدامها في أغلب الأحيان.

- تحليل الشـبكات الاجتماعية: تتبع الروابط بين الأصدقـاء والمعجبين والمتابعين لتحديد روابط التأثير بالإضافة إلى أكبر مصادر التأثير.
- التحليـلات المتقدمـة: تتضمن تحليلات تنبؤيـة وتحليلات النصوص التي تـدرس المحتوى في المحادثات عبر الإنترنت لتحديد المواضيع والمشـاعر والارتباطات التي لن يتم الكشف عنها من خلال المراقبة العرضية.

وتسـتخدم الأدوات والحلـول المتطورة لتحليلات وسـائل التواصل الاجتماعـي جميع فئات التحليلات الثلاث (أي: وصفية، تنبؤية، وقائية) بطريقة تدريجية إلى حد ما.

أفضل الممارسات في تحليلات وسائل التواصل الاجتماعي:

باعتبارها أداة ناشـئة، فإن تحليلات وسـائل التواصل الاجتماعي تُمارس مـن قبل الشركات بطريقة عشـوائية بعض الشيء. ونظرًا لعدم وجود منهجيات راسـخة، فإن الجميع يحاولون إنشـاء منهجياتهم الخاصة عن طريق التجربة والخطأ. وفيما يلي بعض أفضل الممارسات التي تم اختبارها ميدانيًّا لتحليلات وسائل التواصل الاجتماعي التي اقترحها كل من Paine وChaves (2012).

الاعتقـاد بأن القياس نظـام توجيه لا نظام تصويت: غالبًا ما تسـتخدم القياسـات للعقاب أو المكافـآت؛ ولا ينبغـي أن تكون كذلك، بل يجب أن يكونوا على معرفة بالأدوات والممارسـات الأكـثر فاعلية، ومـا يجب إيقافه لأنـه لا يعمل، وما يجـب فعله أكثر لأنه يعمل بشـكل جيد. ونظـام التحليلات الجيد ينبغي أن يخبرك أين تحتاج إلى التركيز. وربما يكون كل هذا التركيز على Facebook غـير مهم لأن ذلك ليس مكان جمهورك، حيـث إنه من الممكن أن يكونوا جميعهم عـلى تويتر أو العكس. ووفقًا لـ Paine وChaves، فإن تفضيل القناة لن يكون بالضرورة بديهيًّا، يقـول كل منهم: "لقد عملنا للتو مع فندق لم يكن لديه أي نشـاط تقريبًا على Twitter لعلامة تجارية واحدة غير أن لديه الكثير من أنشطة Twitter لإحدى علاماتهم التجارية العالية" وبدون أداة قياس دقيقة، فإنك لن تعرف.

تتبع الشـعور: يريد العملاء أن يأخذوا ما يسـمعونه ويتعلمونه من المحادثات عبر الإنترنت ويعملـون عليه. والمفتاح لذلك هو أن تكون دقيقًا في اسـتخراج وتمييز نواياهم من خلال قياس مشاعرهم. وكما رأينا سابقًا في هذا الفصل، فإن الأدوات التحليلية النصية يمكنها تصنيف المحتوى عبر الإنترنت، وكشـف المفاهيم المرتبطة، وكشـف الشـعور في محادثة كـ "إيجابي"، "سلبي"، أو "محايد"، استنادًا إلى الكلمات التي يستخدمها الناس. ومن الناحية المثالية، فإنك سوف ترغب في

أن تكون قادرًا على تمييز المشاعر تجاه منتج معين، وخدمة معينة، ووحدة أعمال محددة. وكلما زادت دقــة فهمك للهجة والإدراك الذي يعبر عنه الناس، أصبحت المعلومات أكثر قابلية للتنفيذ، لأنك تخفف من المخاوف بشــأن التناقض المختلط. ولا يجب وضع علامة "محايدة" على العبارة ذات التناقــض المختلــط، مثل: "فندق في موقع رائع ولكن الحــمام ذو رائحة كريهة" لأن لديك إيجابيات وسلبيات تقابل بعضها البعض. ولكي تكون قابلة للتطبيق، فإنه يجب التعامل مع هذه الأنواع من العبارات بشــكل منفصل. فمثلًا، عندما نقول: "كان الحمام كريه الرائحة" هو شيء يمكن لأي شــخص أن يلمسه بنفسه ويحســن حالته. كما يمكن تقسيم هذه المشاعر وتصنيفها، وإلقــاء نظرة على الاتجاهات بمرور الوقت، ورؤية اختلافــات كبيرة في الطريقة التي يتحدث بها الناس بشكل إيجابي أو سلبي عنك، فضلًا عن مقارنة المشاعر حول علامتك التجارية مع منافسيك.

تحســين دقة النص بشكل مستمر: إن استخدام حزمة تحليل نصوص خاصة بمجال الصناعة سوف يؤدي إلى معرفة مفردات عملك بالفعل. وسيكون للنظام قواعد لغوية مبنية فيه، لكنه يتعلم مع مرور الوقــت ويصبح أفضــل وأفضل. وبقدر تمكنك من ضبط نموذج إحصائي، بقدر حصولك على المزيد من البيانات، أو معلمات شــبكة أفضل، أو تقنيات جديدة لتقديم نتائج أفضل، فإنك ستفعل الشيء نفسه مع معالجة اللغات الطبيعية التي تدخل في تحليل المشــاعر. فإذا قمت بإعداد القواعد، والتصنيفات، بالإضافة إلى التنظيم، ومعاني الكلمات، شاهد كيف تبدو النتائج ثم عُدْ وكررها مرة أخرى.

النظر إلى تأثير تأرجح النتائج: يســعى الجميع للوصول إلى نتيجة رائعة على موقع رفيع المستوى، ولكن هذا ليس كل شيء، فهناك فرق بين النتيجة الرائعة التي تظهر وتختفي في مقابل نتيجة رائعة يتم التغريد بها، وإعادة التغريد، والتقاطها من قبل المدونين المؤثرين. ويجب أن توضح لك التحليلات أي من الأنشطة المتعلقة بوسائل التواصل الاجتماعي تتحول إلى "فيروسية" وأي منها يسكن سريعًا، ولماذا.

النظر إلى ما وراء العلامة التجارية: إن اهتمام الأشخاص بعلامتهم التجارية فقط يعد من أكبر الأخطاء التي يرتكبونها. وللنجاح في تحليل واســتخدام وسائل التواصل الاجتماعي، فيجب عليك تفهم كل ما يثار من القضايا المحيطة بمنتجك أو خدمتك. وعادةً لا يهتم العملاء برســالة الشركة أو علامتها التجارية؛ فهم يهتمون بأنفسهم. لذلك؛ فإنه يجب الانتباه إلى ما يتحدثون عنه، وأين يتحدثون بذلك ومع من.

التعرف على معظم التأثيرات القوية: تعاني المؤسســات مــن صعوبة تحديد صاحب التأثير الأكبر على الرأي العام، وهذا يبين أن أهم المؤثرين ليسوا بالضرورة ممن يؤيدون علامتك التجارية على وجه التحديد؛ ولكنهم يكونون ضمن من يؤثرون على نطاق المحادثة الدائرة حول موضوعك، وينبغي تفهم ما إذا كانوا يعطون آراءً إيجابية، أو يعبرون عن دعمهم، أو ببساطة يقدمون بملاحظات أو نقد، وذلك من خلال فهم طبيعة محادثاتهم، ومكان علامتي التجارية فيما يتعلق بالمنافسة في تلك المنطقة.

النظـرة الثاقبة إلى دقة أدواتـك التحليلية: حتى وقت قريب لم تكن الأدوات الآلية المعتمدة على الحاسـب تتسـم بدرجة الدقة التي يتمتع بها البشر في غربلة محتوى الإنترنت. وحتى الآن تختلف الدقة باختلاف الوسـائل. فبالنسـبة لمواقع مراجعة المنتجـات ومواقع مراجعة الفنادق وتويتر، يمكن أن تصل إلى دقة تتراوح بين ٨٠٪ و٩٠٪ وذلك نظرًا لأن السياق يكون محاصرًا بشكل أكبر. وعندما تبدأ بالنظر إلى المدونات ومنتديات المناقشة، حيث تكون المحادثة واسعة النطاق، يمكن للبرنامج أن يحقق دقة تصل إلى ٦٠٪ (Paine & Chaves, 2012). وسـتزداد هذه الأرقام بمرور الوقت لأن أدوات التحليلات يتم ترقيتها باسـتمرار بقواعد جديدة وخوارزميات محسـنة لتعكس التجربة الميدانية والمنتجات الجديدة وتغير ظروف السوق وأنماط الكلام الناشئة.

دمج ذكاء وسائل التواصل الاجتماعي في التخطيط: بمجرد أن يكون لديك منظور كبير للصورة مـع نظرة تفصيلية، يكون بإمكانك البدء في دمج هذه المعلومات في دورة التخطيط الخاصة بك. غير أن هذا كلام يصعب تنفيذه. وقد كشـف اسـتطلاع سريع للجمهور أن قليل من الناس الآن يقومـون بدمج التعلم من المحادثات عبر الإنترنـت في دورات التخطيط الخاصة بهم (Paine & Chaves, 2012). وتعد إحدى طرق تحقيق ذلك هي إيجاد علاقات زمنية بين مقاييس وسـائل التواصل الاجتماعي وأنشـطة الأعمال الأخرى أو أحداث السوق. وعادةً ما يتم اللجوء إلى وسائل التواصل الاجتماعي إما بشـكل أسـاسي أو بواسطة شيء تقوم به مؤسستك. وبالتالي، إذا لاحظت تصاعدًا في النشاط في أي نقطة من الزمن، فتكون بحاجة لمعرفة أسباب ذلك.

أسئلة مراجعة على القسم ٥-١٠:

١- ما المقصود بالتحليلات الاجتماعية؟ ولماذا تعتبر مهمة؟

٢- ما هي الشبكة الاجتماعية؟ وما هي الحاجة لنظام الحسابات القومية؟

٣- ما هي وسائل التواصل الاجتماعي؟ وكيف ترتبط بـ Web 2.0؟

٤- ما هي تحليلات وسائل التواصل الاجتماعي؟ وما هي أسباب تزايد انتشارها؟

٥- كيف يمكنك قياس تأثير تحليلات وسائل التواصل الاجتماعي؟

تلخيص لأهم نقاط الفصل:

- التنقيب في النص هو اكتشاف المعرفة من مصادر البيانات غير المهيكلة (غالبًا تستند إلى النص). وحيث إن قدرًا كبيرًا من المعلومات يكون في شـكل نـص، فإن التنقيب في النص هو واحد من أسرع فروع مجال ذكاء الأعمال نموًا.

- تتواجد تطبيقات التنقيب في النص في كل مجال من مجالات العمل والحكومة تقريبًا، بما في ذلك التسويق والتمويل والرعاية الصحية والطب والأمن الداخلي.
- يقوم التنقيب في النص باستخدام معالجة اللغة الطبيعية (NLP) للحث على هيكلة مجموعة النصوص، ثم يستخدم بعد ذلك خوارزميات التنقيب في البيانات مثل التصنيف والتجميع والاقتران وتسلسل الاكتشاف لاستخراج المعرفة منها.
- يمكن تعريف الشعور بأنه رأي ثابت يعكس مشاعر المرء.
- يتعامل تحليل المشاعر مع التمييز بين فئتين؛ إيجابية وسلبية.
- كمجال للبحث، يرتبط تحليل المشاعر ارتباطًا وثيقًا بعلم لغات الحاسوب، والبرمجة اللغوية العصبية، والتنقيب في النصوص.
- يحاول تحليل المشاعر الإجابة عن السؤال التالي: "بم يشعر الناس حول موضوع معين؟" عن طريق التعمق في آراء الكثيرين باستخدام مجموعة متنوعة من الأدوات الآلية.
- يعتبر صوت العميل (VOC) جزءًا لا يتجزأ من نظامَي إدارة علاقات العملاء التحليلية وإدارة تجارب العملاء، وغالبًا ما يتم تشغيلها من خلال تحليل المشاعر.
- يدور صوت السوق (VOM) حول فهم الآراء والاتجاهات الكلية على مستوى السوق.
- يتم تحقيق تحديد التناقض في تحليل المشاعر إما باستخدام المعجم كمكتبة مرجعية أو باستخدام مجموعة من وثائق التدريب.
- WordNet هو معجم شهير للأغراض العامة تم إنشاؤه في جامعة برينستون.
- SentiWordNet هو امتداد لـ WordNet لكي يُستخدم لتعريف المشاعر.
- تحليلات الكلام هي مجال متزايد من العلوم يسمح للمستخدمين بتحليل واستخراج المعلومات من المحادثات الحيّة والمُسجلة.
- يمكن تعريف التنقيب في الويب بأنه اكتشاف وتحليل معلومات مثيرة للاهتمام ومفيدة من الويب، وعن الويب، وعادةً ما تستخدم أدوات تستند إلى الويب.
- يمكن النظر إلى التنقيب على الويب على أنه يتكون من ثلاثة مجالات هي: التنقيب في محتوى الويب، والتنقيب في هيكل الويب، والتنقيب في استخدام الويب.
- يشير التنقيب في محتوى الويب إلى الاستخراج التلقائي للمعلومات المفيدة من صفحات الويب، ويمكن استخدامه لتحسين نتائج البحث التي تنتجها محركات البحث.

- يشـير التنقيب في هيكل الويب إلى توليد معلومات مثيرة للاهتمام من الروابط الموجودة ضمن صفحات الويب.
- يمكن اسـتخدام التنقيب في هيكل الويب أيضًـا لتحديد أعضاء مجتمع معين، وربما أيضًا لأدوار الأعضاء في المجتمع.
- يشير تعدين استخدام الويب إلى تطوير معلومات مفيدة من خلال تحليل سجلات خادم الويب، وملفات تعريف المستخدمين، ومعلومات المعاملات.
- ظهر كل من التنقيب في النصوص والتنقيب في الويب كمكونات حيوية للجيل القادم من أدوات ذكاء الأعمال لتمكين المنظمات من التنافس بنجاح.
- محـرك البحث هو برنامج يقوم بالبحث عن الوثائق (مواقع الإنترنت أو الملفات)، اسـتنادًا إلى الكلـمات المفتاحية (كلمات فرديـة، أو مصطلحات متعددة الكلـمات، أو جملة كاملة) التي يقدمها المستخدمون، والتي تتعلق بموضوع استعلاماتهم.
- SEO هو النشاط المعني بالتأثير على رؤية أي موقع تجارة إلكترونية أو أي موقع ويب في نتائج البحث الطبيعية لمحرك البحث (غير المدفوع أو الأساسي).
- VOC هو مصطلح يستخدم عادةً لوصف العملية التحليلية لالتقاط توقعات العميل وتفضيلاته وعدم تفضيلاته.
- التحليـلات الاجتماعيـة هي مراقبـة وتحليل وتفسـير التفاعلات الرقمية وعلاقات الأشـخاص والمواضيع والأفكار والمحتوى.
- الشبكة الاجتماعية هي بنية اجتماعية تتألف من أفراد / أشخاص (أو مجموعات من الأفراد أو المنظمات) ترتبط فيما بينها بنوع من الروابط / العلاقات.
- تشير تحليلات وسائل التواصل الاجتماعي إلى الطرق المنهجية والعلمية لاستهلاك الكمية الهائلة من المحتوى الذي تم إنشـاؤه بواسطة منافذ وأدوات وتقنيات وسائل التواصل الاجتماعي على الويب من أجل تحسين القدرة التنافسية للمنظمات.

مصطلحات أساسية:

اقتران.	معالجة اللغة الطبيعية (NLP).	تحليلات وسائل التواصل الاجتماعي.	البيانات غير المهيكلة.
صفحات موثوقة.	وضع علامات على جزء من الكلام.	شبكة اجتماعية.	صوت العميل (VOC).

تصنيف.	تحديد التناقض.	عنكبوت.	تحليلات الويب.
تحليل النقر.	Polyseme.	الانبثاق.	التنقيب في محتوى الويب
تجمع.	محرك البحث.	كلمات التوقف.	زاحف الويب.
جسم.	تحليل المشاعر.	مصفوفة Term-Document (TDM).	التنقيب في الويب.
كشف الخداع.	SentiWordNet.	التنقيب في النص.	التنقيب في بنية الويب.
المحاور.	تحليل القيمة المفردة (SVD).	Tokenizing.	التنقيب باستخدام الويب.
البحث عن موضوع ناجم رابط تشعبي (HITS).		تحليل النزعة.	WordNet.

أسئلة للمناقشة:

١- اشرح العلاقة بين التنقيب في البيانات، والتنقيب في النص، وتحليل المشاعر.

٢- عرف - بأسلوبك الخاص - التنقيب في النص، وناقش تطبيقاته الأكثر شيوعًا.

٣- ماذا يعني الحث على هيكلة البيانات النصية؟ ناقش الطرق البديلة لهيكلة البيانات النصية.

٤- ما هو دور NLP في التنقيب في النص؟ ناقش قدرات وقيود NLP في سياق التنقيب في النص.

٥- اسرد وناقش ثلاثة مجالات تطبيق بارزة للتنقيب في النص. ما هي الفكرة الرئيسة الشائعة بين مجالات التطبيق الثلاثة التي اخترتها؟

٦- ما هو تحليل المشاعر؟ وكيف يرتبط بالتنقيب في النص؟

٧- ما هي التحديات الشائعة التي يتعين على تحليل المشاعر التعامل معها؟

٨- ما هي مجالات التطبيق الأكثر شيوعًا لتحليل المشاعر؟ ولماذا؟

٩- ما هي الخطوات الرئيسة في تنفيذ مشروعات تحليل المشاعر؟

١٠- ما هما الطريقتان الشائعتان لتحديد التناقض؟ اشرح إجابتك.

١١- ناقش أوجه الاختلاف والشبه بين التنقيب في النص والتنقيب في الويب.

١٢- بأسلوبك الخاص عرِّف التنقيب في الويب، مع مناقشة أهميته.

١٣- ما هي المجالات الرئيسة الثلاثة للتنقيب في الويب؟ ناقش أوجه الاختلاف والشبه بين هذه المجالات الثلاثة.

١٤- ما هو محرك البحث؟ ولماذا يعتبر مهمًّا للشركات؟

١٥- ما هو SEO؟ ومن الذي يستفيد منه؟ وكيف؟

١٦- ما هي تحليلات الويب؟ وما هي المقاييس المستخدمة فيها؟

١٧- عرف كلًا من التحليلات الاجتماعية والشبكة الاجتماعية وتحليلات الشبكة الاجتماعية. وما هي العلاقات فيما بينها؟

١٨- ما هي تحليلات وسائل التواصل الاجتماعي؟ وكيف تتم؟ ومن الذي يقوم بها؟ وماذا يخرج منها؟

تمارين - شبكة جامعة تيراداتا (TUN) وغيرها من التمارين اليدوية:

١- قم بزيارة teradatauniversitynetwork.com. حدد الحالات عن التنقيب في النص. قم بوصف التطورات الأخيرة في هذا المجال. إذا لم تستطع إيجاد حالات كافية على موقع شبكة جامعة تيراداتا على الويب، قم بتوسيع بحثك إلى موارد أخرى تستند إلى الويب.

٢- انتقل إلى teradatauniversitynetwork.com أو حدد مكان الأوراق البيضاء والندوات على الويب والمواد الأخرى المتعلقة بالتنقيب في النصوص. قم بتجميع النتائج الخاصة بك في تقرير قصير مكتوب.

٣- اذهب إلى teradatauniversitynetwork.com وقم بإيجاد دراسة الحالة التي تحمل اسم "eBay Analytics". اقرأ الحالة بعناية، وافهمها جيدًا من خلال البحث في الإنترنت للحصول على معلومات إضافية، والإجابة عن أسئلة الحالة.

٤- انتقل إلى teradatauniversitynetwork.com وابحث عن حالة تحليل المشاعر التي تحمل اسم "How Do We Fix an App Like That?" اقرأ الوصف، واتبع الإرشادات لتنزيل بيانات وأداة تنفيذ التمرين.

٥- قــم بزيــارة teradatauniversitynetwork.com. حدد الحالات عــن التنقيب في الويب. قم بوصف التطورات الأخيرة في هذا المجال. إذا لم تســتطع إيجاد حالات كافية على موقع شبكة جامعة تيراداتا على الويب، قم بتوسيع بحثك إلى موارد أخرى تستند إلى الويب.

٦- تصفح الويب وقواعد البيانات الرقمية الخاصة بمكتبتك لتحديد المقالات التي تصنع الرابط بين التنقيب في النصوص / الويب والأنظمة المعاصرة لذكاء الأعمال.

مشروعات مهام الفريق ولعب الأدوار:

١- افحص كيفية التقاط البيانات النصية تلقائيًّا باستخدام تقنيات الويب. وفور التقاطها، ما هي الأنماط المحتملة التي يمكنك استخراجها من هذه المصادر غير المهيكلة للبيانات؟

٢- قم بمقابلة المســؤولين في كليتك أو المســؤولين التنفيذيين في مؤسستك لتحديد كيف يمكن أن يساعدهم التنقيب في النصوص والتنقيب في الويب في عملهم. اكتب اقتراحًا يصف نتائجك، ثم أعدَّ تقريرًا يتضمن تقديرات التكلفة والفوائد.

٣- انتقــل إلى مــوارد مكتبتك على الإنترنــت. تعلم كيفية تنزيل ســمات مجموعة من الأدبيات (مقالات الدوريات) في موضوع محدد، ثم قم بتنزيل البيانات ومعالجتها باســتخدام منهجية مشابهة لتلك التي تم شرحها في الحالة العملية (٥-٥).

٤- ابحث عن مجموعة بيانات نص المشاعر المتاحة بسهولة (انظر الرؤية الفنية (٥-٣)للحصول على قائمة بمجموعات البيانات الشائعة) وقم بتنزيلها إلى جهاز الحاسب الخاص بك. إن كان لديك أداة تحليليــة قادرة على التنقيب في النصوص، فاســتخدمها. وإن لم يكن، قم بتنزيل وتثبيت RapidMiner (http://rapid-i.com). قم أيضًا بتثبيت الوظيفة الإضافية Text Analytics لبرنامج RapidMiner. قم بمعالجة البيانات التي تم تنزيلها باستخدام أداة التنقيب في النص (أي تحويل البيانات إلى نموذج منظم)، ثم قم ببناء النماذج وتقييم دقة الكشــف عن المشاعر في العديد من نماذج التصنيف (مثل: آلة المتجهات الداعمة، أشجار القرار، الشبكات العصبية، الانحدار اللوجستي). اكتب تقريرًا تفصيليًّا تشرح فيه نتائجك وخبراتك.

٥- افحص كيفية التقاط البيانات المســتندة إلى الويب تلقائيًّا باســتخدام أحدث التقنيات. وفور التقاطها، ما هي الأنماط المحتملة التي يمكنك استخراجها من هذه المصادر الغنية بالمحتوى للبيانات، والتي تكون في معظمها غير منظمة؟

تمارين الإنترنت:

١- قـم بإيجاد حالات حديثة لتطبيقـات ناجحة عن التنقيب في النصـوص والتنقيب في الويب. جرِّب موردي وشركات الاستشارات لبرامج التنقيب في النصوص والويب وابحث عن حالات أو قصص نجاح، ثم قم بإعداد تقرير يلخص خمس دراسات حالة جديدة.

٢- انتقـل إلى statsoft.com. اختر Downloads، وقـم بتنزيل ثلاث صفحات بيضاء على الأقل حـول التطبيقات. أي من هذه التطبيقات اسـتخدم تقنيات التنقيـب في البيانات / النص / الويب التي تمت مناقشتها في هذا الفصل؟

٣- انتقـل إلى sas.com، وقـم بتنزيل ثلاث صفحات بيضاء عـلى الأقل حول التطبيقات. أي من هذه التطبيقات اسـتخدم تقنيات التنقيب في البيانات / النص / الويب التي تمت مناقشـتها في هذا الفصل؟

٤- انتقـل إلى ibm.com، وقم بتنزيل ثـلاث صفحات بيضاء على الأقل حول التطبيقات. أي من هذه التطبيقات اسـتخدم تقنيات التنقيب في البيانات / النص / الويب التي تمت مناقشـتها في هذا الفصل؟

٥- انتقـل إلى teradata.com، وقم بتنزيل ثلاث صفحات بيضاء على الأقل حول التطبيقات. أي من هذه التطبيقات استخدم تقنيات التنقيب في البيانات / النص / الويب التي تمت مناقشتها في هذا الفصل؟

٦- انتقـل إلى clarabridge.com، وقم بتنزيل ثلاث صفحات بيضاء على الأقل حول التطبيقات. أي من هذه التطبيقات استخدم تقنيات التنقيب في النص بطريقة مبتكرة؟

٧- انتقل إلى kdnuggets.com، قم باستكشـاف الأقسـام حول التطبيقـات وكذلك البرامج. قم بإيجاد أسماء ثلاث حزم إضافية على الأقل للتنقيب في البيانات والتنقيب في النصوص.

٨- قم بمسـح بعض أدوات وموردي التنقيـب في الويب. حدد بعض منتجات التنقيب في الويب ومقدمي الخدمات غير المذكورة في هذا الفصل.

٩- انتقـل إلى attensity.com، وقم بتنزيل ثلاث صفحات بيضاء على الأقل حول التطبيقات. أي من هذه التطبيقات استخدم تقنيات التنقيب في البيانات / النص / الويب التي تمت مناقشتها في هذا الفصل؟

المراجع:

- Bond C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. Personality and Social Psychology Reports, 10(3), 214-234.
- Brogan, C., & Bastone, J. (2011). Acting on customer intelligence from social media: The new edge for building cus- tomer loyalty and your brand. SAS white paper.
- Chun, H. W., Tsuruoka, Y., Kim, J. D., Shiba, R., Nagata, N., & Hishiki, T. (2006). Extraction of gene-disease relations from MEDLINE using domain dictionaries and machine learning. Proceedings of the 11th Pacific Symposium on Biocomputing, 4-15.
- Coussement, K., & Van Den Poel, D. (2008). Improving customer complaint management by automatic email classification using linguistic style features as predictors. Decision Support Systems, 44(4), 870-882.
- Coussement, K., & Van Den Poel, D. (2009). Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. Expert Systems with Applications, 36(3), 6127-6134.
- Cutts, M. (2006, February 4). Ramping Up on International Webspam. mattcutts.com/blog. mattcutts.com/blog/ramp ing-up-on international-webspam (accessed March 2013). Delen, D., & Crossland, M. (2008). Seeding the survey and analysis of research literature with text mining. Expert Systems with Applications, 34(3), 1707-1720.
- Esuli, A., & Sebastiani, F. (2006, May). SentiWordNet: A publicly available lexical resource for opinion mining. In Proceedings of LREC, 6, 417-422.
- Etzioni, O. (1996). The World Wide Web: Quagmire or gold mine? Communications of the ACM, 39(11), 65-68.
- EUROPOL. (2007). EUROPOL Work Program 2005. statewatch. org/news/2006/apr/europol-work-programme-2005.pdf (accessed October 2008).
- Feldman, R., & Sanger, J. (2007). The text mining handbook: Advanced approaches in analyzing unstructured data. Boston: ABS Ventures.
- Fuller, C. M., Biros, D., & Delen, D. (2008). Exploration of feature selection and advanced classification models for high-stakes deception detection. Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS), Big Island, HI: IEEE Press, 80-99.
- Ghani, R., Probst, K., Liu, Y., Krema, M., & Fano, A. (2006). Text mining for product attribute extraction. SIGKDD Explorations, 8(1), 41-48.
- Goodman, A. (2005). Search engine showdown: Black hats versus white hats at SES. SearchEngineWatch. searchenginewatch.com/article/2066090/Search-Engine-Showdown-Black-Hats-vs.-White-Hats-at-SES (accessed February 2013).

- Han, J., & Kamber, M. (2006). Data mining: Concepts and techniques, 2nd ed. San Francisco: Morgan Kaufmann.
- Harvard Business Review. (2010). The new conversation: Taking social media from talk to action. A SAS-Sponsored Research Report by Harvard Business Review Analytic Services. sas.com/resources/whitepaper/wp_23348.pdf (accessed March 2013).
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. Business Horizons, 53(1), 59-68.
- Kim, S. M., & Hovy, E. (2004, August). Determining the sentiment of opinions. In Proceedings of the 20th International Conference on Computational Linguistics (p. 1367). Association for Computational Linguistics.
- Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5), 604-632.
- Lin, J., & Demner-Fushman, D. (2005). "Bag of words" is not enough for strength of evidence classification. AMIA Annual Symposium Proceedings, 1031-1032. pubmedcentral.nih.gov/articlerender.fcgi?artid=1560897.
- Liu, B., Hu, M., & Cheng, J. (2005, May). Opinion observer: Analyzing and comparing opinions on the Web. In Proceedings of the 14th International Conference on World Wide Web (pp. 342-351). ACM.
- Mahgoub, H., Rösner, D., Ismail, N., & Torkey, F. (2008). A text mining technique using association rules extraction. International Journal of Computational Intelligence, 4(1), 21-28.
- Manning, C. D., & Schutze, H. (1999). Foundations of statistical natural language processing. Cambridge, MA: MIT Press.
- McKnight, W. (2005, January 1). Text data mining in business intelligence. Information Management Magazine. informationmanagement.com/issues/20050101/10164871.html (accessed May 22, 2009).
- Mejova, Y. (2009). Sentiment analysis: An overview. Comprehensive exam paper. http://www.cs.uiowa.edu/~ymejova/publications/CompsYelenaMejova.pdf (accessed February 2013).
- Miller, T. W. (2005). Data and text mining: A business applications approach. Upper Saddle River, NJ: Prentice Hall.
- Morgan, N., Jones, G., & Hodges, A. (2010). The complete guide to social media from the social media guys. thesocialmediaguys.co.uk/wpcontent/uploads/downloads/2011/03/CompleteGuidetoSocialMedia.pdf (accessed February 2013).
- Nakov, P., Schwartz, A., Wolf, B., & Hearst, M. A. (2005). Supporting annotation layers for natural language processing. Proceedings of the ACL, Interactive Poster and Demonstration Sessions, Ann Arbor, MI. Association for Computational Linguistics, 65-68.
- Paine, K. D., & Chaves, M. (2012). Social media metrics. SAS white paper. sas.com/resources/whitepaper/wp_19861. pdf (accessed February 2013).

- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Hanover, MA: Now Publishers, available at http://books.google.com.
- Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009, August). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 (pp. 248-256). Association for Computational Linguistics.
- Schmidt, L.-H. (1996). Commonness across cultures. In A. N. Balslev (Ed.), Cross-cultural conversation: Initiation (pp. 119-132). New York: Oxford University Press.
- Scott, W. R., &. Davis, G. F. (2003). Networks in and around organizations. In Organizations and Organizing. Upper Saddle River: NJ: Pearson Prentice Hall.
- Shatkay, H., Höglund, A., Brady, S., Blum, T., Dönnes, P., & Kohlbacher, O. (2007). SherLoc: High-accuracy prediction of protein subcellular localization by integrating text and protein sequence data. Bioinformatics, 23(11), 1410-1415.
- Snyder, B., & Barzilay, R. (2007, April). Multiple aspect ranking using the good grief algorithm. In HLT-NAACL (pp. 300-307).
- Strapparava, C., & Valitutti, A. (2004, May). WordNet affect: An affective extension of WordNet. In LREC (Vol. 4, pp. 1083-1086).
- The Westover Group. (2013). 20 key Web analytics metrics and how to use them. http://www.thewestovergroup.com (accessed February 2013).
- Thomas, M., Pang, B., & Lee, L. (2006, July). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (pp. 327-335). Association for Computational Linguistics.
- Weng, S. S., & Liu, C. K. (2004). Using text classification and multiple concepts to answer e-mails. Expert Systems with Applications, 26(4), 529-543.

الفصل السادس

التحليلات التوجيهية: التحسين والمحاكاة

أهداف التعلم:

- فهم تطبيقات تقنيات التحليلات التوجيهية بالاشتراك مع إعداد التقارير والتحليلات التنبؤية.
- استيعاب المفاهيم الأساسية لنمذجة القرارات التحليلية.
- استيعاب مفاهيم النماذج التحليلية لمشاكل القرار المختارة، بما في ذلك البرمجة الخطية ونماذج المحاكاة لدعم القرار.
- وصف كيفية استخدام جداول البيانات بالنسبة لكل من النمذجة التحليلية والحلول التحليلية.
- شرح المفاهيم الأساسية للتحسين وتوقيت استخدامها.
- وصف كيفية هيكلة نموذج برمجة خطية.
- شرح المقصود بتحليل الحساسية، وتحليل «ماذا - لو»، والبحث عن الهدف.
- استيعاب مفاهيم وتطبيقات أنواع المحاكاة المختلفة.
- فهم التطبيقات المحتملة لمحاكاة الأحداث المنفصلة.

يعد هذا الفصل إضافة جديدة إلى هذا الكتاب، حيث يتخطى مرحلة تطبيقات التحليلات وصولًا إلى ما بعد إعداد التقارير والتحليلات التنبؤية. ويغطي هذا الفصل التقنيات المختارة التي يمكن توظيفها جنبًا إلى جنب مع النماذج التنبؤية للمساعدة في دعم القرار. ونركز هنا على التقنيات التي يسهل تنفيذها نسبيًا سواء باستخدام أدوات جداول البيانات أو باستخدام أدوات برمجية قائمة بذاتها. وبالطبع، يوجد الكثير من التفاصيل الإضافية ليتم تعلمها حول نماذج علوم الإدارة، ولكن الهدف من هذا الفصل هو فقط توضيح ما هو ممكن وكيفية تنفيذه ضمن إعدادات حقيقية.

ونعرض فيما يلي تقديمًا لهذه المادة مع ملاحظة الحذر: قد تشكل النمذجة موضوعًا صعبًا، كما تعد فنًا بجانب كونها علمًا. وليس بالضرورة أن تكون الغاية من هذا الفصل أن تصبح خبيرًا بالموضوعات التي تتعلق بالنمذجة والتحليل، ولكن تُستخدَم المادة لاكتساب المعرفة بالمفاهيم المهمة من حيث صلتها بالتحليلات التوجيهية واستخدامها في صنع القرار. ومن المهم أن ندرك أن النمذجة التي نحن بصدد الحديث عنها هنا ترتبط فقط بشكل سريع بمفاهيم نمذجة البيانات،

فيجب عدم الخلط بينهما. إننا نتحرك عبر بعض المفاهيم والتعاريف الرئيسة لنمذجة القرار، ونقدم بعد ذلك فكرة النمذجة بشكل مباشر في جداول البيانات، ثم بعد ذلك نناقش الهيكل والتطبيق لنماذج ومنهجيات برهنت على نجاحها عبر الزمن: البرمجة الخطية ومحاكاة الحدث المنفصل. وكما ذُكِرَ سابقًا، من الممكن الحصول على دورات متعددة - تحديدًا - في هذين الموضوعين، ولكن هدفنا هو منحك إحساس بما هو ممكن، وذلك من خلال الأقسام المتنوعة لهذا الفصل.

٦-١ مقال افتتاحي: منطقة فيلاديلفيا التعليمية تستخدم التحليلات التوجيهية لإيجاد حلٍ أمثل لإبرام عقود خطوط سير الحافلات:

خلفية:

يعد اختيار أفضل الموردين للعمل مهمة مرهقة بل وذات أهمية كبيرة للشركات والمنظمات الحكومية. فبعد تقديم كل مورد لعطائه بخصوص مهمة معينة من خلال عملية المناقصة، تقوم الشركة أو المنظمة بتقييم كل اقتراح ومن ثم اتخاذ قرار بشأن اختيار المورد الأنسب لاحتياجاتهم. وعادةً، تضطر الحكومات لاستخدام عملية تقديم العطاءات للمفاضلة بين الموردين ومن ثم الاختيار من بينهم. كانت منطقة فيلاديلفيا التعليمية تسعى للاستعانة بمصادر خارجية لبعض خطوط الحافلات الخاصة بهم؛ إذ لا تمتلك المقاطعة إلا عددًا قليلًا من حافلات المدارس، لا يكفي لخدمة الطلاب القاطنين بها، فكانوا يخدمون بحافلاتهم الخاصة من ٣٠٪ إلى ٤٠٪ فقط من الطرق، فقرروا الاستعانة بموردي خدمة الحافلات الخاصة كمصدر خارجي لتغطية بقية الطرق. تم تكليف تشارلز لويتز (Charles Lowitz)، المنسق المالي لمكتب النقل، بتحديد كيفية تعظيم العائد على الاستثمار وتحسين طريقة إبرام العقود مع مختلف الموردين.

في الماضي، كانت عملية اختيار موردي خدمة الحافلات الذين يتم التعاقد معهم وفقًا للقيود المالية والزمنية عملية مرهقة إذ كانت تتم يدويًا. بالإضافة إلى ذلك، كانت المتغيرات والعوامل المختلفة التي يجب مراعاتها تزيد الأمر تعقيدًا. فقد كانت عملية تقييم الموردين لهذه الخدمة تتم بناءً على خمسة متغيرات وهي: التكلفة، والقدرات، والأمانة، والاستقرار المالي، والفراسة. قام كل مورد بتقديم عطاءه بسعر مختلف للطرق المختلفة. وضع بعض الموردين حدًا أدنى لعدد خطوط السير، وفي حالة عدم مقابلة هذا العدد للمطلوب، فستزيد التكلفة. وكان Lowitz بحاجة لمعرفة كيفية دمج المعلومات لكل عطاء تم تقديمه، وذلك لتوزيع خطوط السير على الموردين بحيث تتم تغطية كل الطرق بأقل تكلفة.

الحل:

بحث Lowitz في البداية عن برنامج يستطيع استخدامه جنبًا إلى جنب مع نموذج العقد الخاص به على Excel. فبدأ باستخدام برنامج Premium Solver من شركة Frontline Systems، مما مكنه من تحديد أفضل الموردين ماليًا وتشغيليًا، وقام بإنشاء نموذج أمثل يأخذ في الحسبان المتغيرات الخمسة سالفة الذكر المتعلقة بكل مورد. وقد تضمن النموذج متغيرات ثنائية صحيحة (نعم / لا) لكل خط من خطوط السير ليتم منحها للموردين الذين تقدموا لتغطية خط سير معين بتكلفة معينة. وقد بلغت هذه المتغيرات ذات (نعم / لا) ما يقرب من ١٦٠٠ متغير. كما تضمن النموذج أيضًا قيودًا تشير إلى أنه سيتم تخصيص مورد واحد فقط لكل خط سير، وبالطبع لا بد من تغطية كل خطوط السير. وتم وضع حد أدنى لعدد خطوط السير التي يخدمها أي مورد، كما تم وضع قيود لبعض التفاصيل الأخرى. ويمكن كتابة كل هذه القيود كمعادلات وإدخالها في نموذج برمجة خطية تامة. وتجدر الإشارة إلى أن هناك العديد من أدوات البرامج تستطيع صياغة مثل هذه النماذج وحلها، غير أن استخدام Microsoft Excel يجعل فهم النموذج أمرًا أكثر سهولة. فتم تضمين برنامج Solver الخاص بشركة Frontline Systems داخل أدوات Microsoft Excel للقيام بحل المشاكل الصغيرة مجانًا. أما النماذج الأكبر حجمًا والأكثر تعقيدًا فيمكن حلها عن طريق شراء إصدار أكبر. وهذا ما فعله Lowitz.

الفوائد:

ساعد النموذج في تطوير حجم كل عقد من العقود، فضلًا عن تحديد عدد الموردين الواجب التعاقد معهم. حيث يتفاوت حجم العقود من مورد لآخر وذلك ما بين ٤-٩٧ خط سير للمورد الواحد. وفي النهاية، تمكنت منطقة فيلادیلفیا التعليمية من وضع خطة تحتوي على العدد الأمثل لشركات توريد الحافلات باستخدام برنامج Excel بدلًا من العملية اليدوية. وهكذا تمكنت المقاطعة من توفير الوقت والمال وذلك باستخدام الأدوات التحليلية لبرنامج Premium Solver لإنشاء نموذج أمثل بالمتغيرات المختلفة.

ماذا يمكننا أن نتعلم من هذه المقالة القصيرة:

تواجه معظم المنظمات مشكلة صنع القرارات حين يلزم اختيار واحد من عدة خيارات. إذ يكون لكل خيار تكلفة وإمكانية يرتبطان به. وتهدف مثل هذه النماذج إلى اختيار توليفة من الخيارات تلبي جميع المتطلبات بجانب تدنية التكاليف. كما يجب ملاحظة أنه يتم تطبيق

التحليلات التوجيهية بشكل خاص على مثل هذه المشكلة المتعلقة بالقرارات. ووجود أدوات مثل Premium Solver المتاح داخل Excel يجعل من السهولة بمكان تطبيق مثل هذه التقنيات.

أسئلة مراجعة على المقالة الافتتاحية:

١- ما هو القرار المتخذ في هذه المقالة الإفتتاحية؟

٢- ما هي البيانات (الوصفية أو التنبؤية) التي قد يتم الاحتياج إليها لعمل أفضل التوزيعات في هذا السيناريو؟

٣- ما هي التكاليف أو القيود الأخرى الواجب مراعاتها عند إبرام مثل هذه العقود المتعلقة بخطوط السير؟

٤- ما هي المواقف الأخرى التي يمكن تطبيق مثل هذه النماذج عليها؟

٦-٢ اتخاذ القرار بناءً على النموذج:

إن اتخاذ القرارات باستخدام أحد أنواع النموذج التحليلي هو ما نسميه التحليلات التوجيهية، وذلك كما تبين من المقالة القصيرة السابقة. وقد تعلمنا في العديد من الفصول السابقة قيمة وعملية معرفة تاريخ ما يجري واستخدام هذه المعلومات أيضًا من أجل التنبؤ بالأمور المُرجح حدوثها. وعلى أية حال، فإننا سنحاول من خلال هذا التمرين تحديد ما ينبغي عمله فيما هو آتٍ. وقد يترتب على ذلك تحديد العملاء المُرجح قيامهم بالشراء منا وتقديم عرض أو إعطاء سعر يزيد من احتمالية قيامهم بالشراء، وبالتالي فإن ذلك سيؤدي إلى تعظيم الأرباح إلى الحد الأمثل. وعلى العكس من ذلك، فقد ينطوي الأمر على القدرة على التنبؤ بالعميل الذي يحتمل اتجاهه إلى مكان آخر والقيام بتقديم عرض مُغري له للاحتفاظ به كأحد العملاء ورفع قيمتنا لديه. وللتأكد من تغطية جميع احتياجاتنا وتقليل التكلفة إلى أدنى حد ممكن فإننا قد نحتاج إلى اتخاذ بعض القرارات بخصوص منح العقود للموردين. ومن الممكن أن نواجه وضعًا ما يتطلب تحديد أي العملاء المحتملين ينبغي حصوله على مواد الحملة الترويجية بحيث لا تكون تكاليف الترويج التي نتحملها باهظة، ونقوم بزيادة معدل الاستجابة إلى أقصى حد ممكن بما لا يتعدى حدود الميزانية. وقد نتخذ قرارًا بشأن ما يجب دفعه مقابل كلمات البحث الرئيسة غير المجانية لرفع عائد الاستثمار لميزانية الإعلانات الخاصة بنا إلى أقصى حد ممكن. وعلى صعيد آخر، فإننا قد نضطر إلى دراسة تاريخ أنماط وصول عملائنا واستخدام تلك المعلومات من أجل التنبؤ بمعدلات الوصول المستقبلية، وتطبيق ذلك لجدولة عدد مناسب من موظفي المتجر لرفع استجابات العملاء إلى أقصى حد بالإضافة إلى تحسين تكاليف العمالة لدينا. ويمكننا من خلال الاعتماد على تحليلاتنا

Source: Adapted with permission from "Optimizing Vendor Contract Awards Gets an A+," http:// www.solver.com/news/optimizing-vendor-contract-awards-gets, 2016 (accessed July 2016).

وتوقعاتنا للطلب على منتجاتنا وتكاليف سلسـلة التوريد، اتخاذ قرارات بشأن المواقع التي سنقوم فيها بإنشاء مستودعاتنا. بالإضافة إلى أن المعلومات الخاصة بأحجام المنتجات التي سيتم تسليمها في مواقع مختلفة وتكاليف التسليم وتوافر سيارة للنقل تسمح لنا بتحديد طرق التسليم اليومية. ويستطيع الشخص الحصول على مئات الأمثلة الخاصة بالحالات التي تكون فيها القرارات المستندة إلى البيانـات ذات قيمـة. وفي الواقع، فإن القدرة على اسـتخدام رؤى وصفية وتنبؤية لمسـاعدة صانـع القـرار على اتخاذ قرارات أفضل تُعـد أكبر فرصة متاحة لمهنة التحليـلات المتنامية. وعلى الرغـم مـن وجود حالات يتم اتخاذ القـرارات فيها من خلال الاعتماد عـلى الخبرة والحدس، غير أن الأرجـح أن القرارات التي يدعمها نموذج ما ستسـاعد صانع القرار عـلى اتخاذ قرارات أفضل. وبالإضافة إلى ذلك أيضًا، فإن هذه القرارات التي يدعمها نموذج ما سـتوفر لصانعي القرار مبررات للتوصيات التي يقدمونها. وهكذا برزت التحليلات التوجيهية باعتبارها أقصى ما انتهى إليه العلم في مجال التحليلات ليتم تطبيقه لاحقًا، وهي تنطوي بشـكل أسـاسي على استخدام نموذج تحليلي للمسـاعدة في توجيه صانع القرار في اتخاذ القرار، أو جعـل عملية اتخاذ القرار عملية آلية وذلك حتـى يتمكن النموذج من تقديم توصيات أو قـرارات. ونظرًا لأن التحليلات التوجيهية تركز على تقديم التوصيات أو اتخاذ القرارات، فإن البعض يُسمي هذه الفئة من التحليلات بتحليلات القرار.

وتتضمن جميع منشورات INFORMS مثل Interfaces وORMS TODAY ومجلة التحليلات Analytics، القصـص التـي توضـح التطبيقات الناجحة لنـماذج القرار في الإعـدادات الحقيقية. ويشتمل هذا الفصل على العديد من الأمثلة على مثل هذه التطبيقات التحليلية الإرشادية. وتجدر بنا الإشارة إلى أن تطبيق نماذج لحالات واقعية يمكنها أن توفر ملايين الدولارات أو تدر عائدًا بملايين الدولارات. وقد وصف كل من Christiansen وآخرون (٢٠٠٩) تطبيقات هذه النماذج في عمليات شركة الشـحن باستخدام TurboRouter، بأنها نظام دعم القرار (DSS) لتوجيه السفن والجدولة، وهم يدعون أن الشركة اسـتخدمت هذا النموذج على مدى ثلاثة أسـابيع لتحسـين الاستفادة من أسـطولها، الأمـر الذي أدى إلى تحقيق ربح إضافي يتراوح من مليـون إلى مليوني دولار في مثل هذه الفترة الزمنية القصيرة. وفي الحالة العملية (٦-١) نقدم مثالًا آخر لتطبيق نموذج.

حالة عملية ٦-١

النقل الأمثل للبيانات من الخادم Downstream لشركة إكسون موبيل ExxonMobil من خلال نظام دعم القرار DSS

شركة إكسـون موبيل هي شركة بترول وغاز طبيعـي، وتعمل في العديد من البلدان في جميـع أنحاء العـالم، وتوفر الشركة المنتجـات البترولية المختلفة، بمـا في ذلك الوقود

النظيـف، ومواد التشـحيم، والمنتجات ذات القيمـة العالية بالإضافـة إلى المواد الخام للعديد من العملاء. ويتم اسـتكمال ذلك من خلال سلسـلة توريد معقدة بين مصافيها وعملائها. ويُعد زيت غاز التفريغ vacuum (VGO) هو أحد المنتجات الرئيسة لشركة ExxonMobil الناقلة transports. وتقوم شركة إكسـون موبيل ExxonMobil بنقل العديد من السفن المُحملة بشحنات shiploads زيت غاز التفريغ VGO من أوروبا إلى الولايات المتحدة. وتشـير التقديرات إلى أن إكسون موبيل تقوم بنقل حوالي من ٦٠-٧٠ سفينة محملة بزيت غاز التفريغ VGO عبر المحيط الأطلسي، في العام الواحد. وقد كان مـن المقرر حتى الآن أن تتم عملية نقل زيت غـاز التفريغ VGO عبر المحيط الأطلسي من خلال عملية نقل يدوية مرهقة سـواء تم هذا النقل من خلال السـفن التي تديرها شركة إكسون موبيل أو سفن طرف ثالث third-party على حد سواء. وتتطلب العملية برمتها تعاون العديد من الافراد عبر منظمة سلسـلة التوريد. ولتحديد الجدول الزمني للنقل الخاص بالسـفن فإنه يتم اسـتخدام العديد من جـداول البيانات المخصصة ذات القيود والمتطلبات الخاصة والمقايضات الاقتصادية. وقد شملت بعض القيود ما يلي:

١- تفاوت توقعات الإنتاج والطلب بصفة دائمة.

٢- قيود الحد الأدنى والحد الأقصى من المخزون.

٣- مجموعة من السفن المتنوعة (مثل: السفن ذات السرعات المتفاوتة وحجم البضائع).

٤- السفن التي يتم تنزيلها وتفريغها في منافذ متعددة.

٥- كل من إكسون موبيل المُدارة ومستلزمات الطرف الثالث والمنافذ.

٦- تكلفة النقل المعقدة والتي تتضمن تغطية زيادة التكاليف وغرامات التأخير.

٧- حجم السفينة وحدود المسارات للموانئ المختلفة.

ويتعذر على العملية اليدوية القيام بتحديد المسارات الفعلية للسفن، وتوقيت كل سـفينة، وكمية زيت غاز التفريغ VGO المحملـة أو المفرغة. وبالإضافة إلى ذلك، فإن بيانات الإنتاج والاستهلاك في عدة مواقع جعلت العملية اليدوية مرهقة وغير فعالة.

المنهجية / الحل:

وقـد تم تطوير أداة دعم القرار التي تسـاعد المُجدولين عـلى التخطيط لجدول زمني مثالي للسفن لتنزيل، ونقل، وتفريغ زيت غاز التفريغ VGO من وإلى عدة مواقع. وقد تم صياغة المشـكلة كمشـكلة البرمجة الخطية العددية. وقد كان من الضروري أن يلبي الحل

المتطلبات الخاصة بالتوجيه، والنقل، والجدولة، وإدارة المخزون إزاء تباينات الإنتاج والطلب المختلفة. وقد اســتُخدِمَت لغة البرمجة الرياضية GAMS لصياغة المشــكلة، كما استُخدِم مايكروسوفت إكسل Microsoft Excel كواجهة المستخدم. وعندما يتم تشغيل آلة الحل (ILOG CPLEX) فإنه يتم التوصل إلى حل أمثل في نقطة تتوقف فيها القيمة الموضوعية للحل الحالي عن التحسن، حيث يتم تحديد معيار التوقف خلال كل برنامج تشغيل.

النتائج / الفوائد:

ومن المتوقع أن يؤدي استخدام نموذج التحسين إلى خفض تكاليف الشحن بالإضافة إلى خفض النفقات الخاصة بغرامات التأخير. ونظرًا لأن الأداة ســتكون قادرة على دعم اســتخدام الســفن بطريقة أفضل كما أنها تســاعد في اختيار السفن (مثل: Panamax مقابــل Afamax) وتصميم جداول التوجيه الأمثل، فإن ذلك من شــأنه تحقيق هدف خفض تكاليف الشــحن بالإضافة إلى خفض النفقــات الخاصة بغرامات التأخير. وينوي الباحثون مواصلة البحث عن طريق استكشــاف أســاليب رياضية أخرى بديلة من أجل حل مشكلة الجدولة. كما أنهم يعتزمون أيضًا منح أداة أنظمة دعم القرار القدرةَ على النظر في منتجات متعددة لمجموعة من السفن.

أسئلة للمناقشة:

١- اذكر ثلاث طرق توضح كيف يمكن أن تؤدي الجدولة اليدوية للسفن إلى تحمُّل المزيد من التكاليف التشغيلية مقارنةً بالأداة التي تم تطويرها.

٢- كيف يمكن لشركة إكســون موبيل الاستفادة بطرق أخرى من أدوات دعم القرار التي تم تطويرها من أجل توسيع وتحسين عملياتها التجارية الأخرى؟

٣- مــا هي بعض القرارات الإســتراتيجية التي يمكن اتخاذها مــن قِبل صناع القرار باستخدام الأداة التي تم تطويرها؟

Source: Adapted from Furman, K. C., Song, J. H., Kocis, G. R., McDonald, M. K., & Warrick, P. H. (2011). Feedstock routing in the ExxonMobil downstream sector. Interfaces, 41(2), 149163-.

أمثلة نموذج تحليلات توجيهية:

تعتمــد التحليلات التوجيهية على النمذجة كعنصر أســاسي. وفي الأمثلــة التي ذكرناها آنفًا في حالات التقديم والتطبيق ينبغي على الفرد استخدام نموذج رياضي حتى يتمكن من التوصية بقرار

لأي مشكلة واقعية. فعلى سبيل المثال، نجد أن تحديد العميل (من بين ملايين العملاء المحتملين) الذي سوف يتلقى عرضًا بغرض تعظيم قيمة الاستجابة الإجمالية مع التقيد بحدود الميزانية، ليس بالأمـر الـذي يمكنك القيام به يدويًا. وبناء نموذج زيادة الاستجابة إلى الحد الأقصى المستند إلى الاحتمالية مع التقيد بحدود الميزانية، من شأنه أن يوفر لنا المعلومات التي ننشدها. واستنادًا إلى المشكلة التي نتناولها، فإن هناك فئات عديدة من النماذج، بالإضافة إلى أنه غالبًا ما يكون هناك العديد من التقنيات المتخصصة لحل كل مشكلة منها. وفي هذا الفصل سوف نتعرف على اثنين من أساليب النمذجة المختلفة. ومن الجدير بالذكر أن معظم الجامعات لديها العديد من الدورات التدريبية التي تغطي هذه الموضوعات تحت عناوين مختلفة مثل بحوث العمليات، وعلوم الإدارة، ونُظم دعم القرار، والمحاكاة التي يمكنها مساعدتك في بناء المزيد من الخبرات في هذه المواضيع. ونظرًا لأن التحليلات التوجيهية تنطوي عادةً على تطبيق نماذج رياضية، فإن مصطلح علم البيانات أحيانًا يرتبط بتطبيق مثل هذه النماذج الرياضية بشكل أكثر انتشارًا. وقبل أن نتعرف على دعم النمذجة الرياضية في التحليلات التوجيهية، دعونا نفهم أولًا بعض مشاكل النماذج.

تحديد المشكلة والتحليل البيئي:

لا يتـم اتخاذ القرارات من فراغ بل لابد من تحليـل نطاق المجال وقوى وديناميكيات البيئة، حيـث يحتاج صانع القرار إلى تحديد الثقافة التنظيميـة وعمليات صنع القرار في الشركات (على سـبيل المثال: من يقوم باتخاذ القرارات، ودرجة المركزية). فمن الوارد أن يكون ظهور المشـكلة الحالية سـببه العوامل البيئية، وهذا قد يسمى بشكل رسمي المسح والتحليل البيئي، ويقصد به رصد المعلومات المجمعة ومسـحها وتفسـيرها. ويمكن لأدوات ذكاء الأعمال / تحليلات الأعمال (BI/BA) أن تسـاعد في تحديد المشـاكل عن طريق القيام بعملية مسـح لها. وينبغي أن تكون المشـكلة مفهومة جيدًا، كما يجب أن يشـارك الجميع نفس إطار الفهم نظرًا لأن المشكلة سوف يتم تمثيلها في النهاية بالنموذج بشكل أو بآخر. وبغير ذلك، فإن النموذج لن يساعد صانع القرار.

تحديد المتغير: يُعد تحديد متغيرات النموذج (مثل: القرار، والنتيجة، وعدم التحكم فيهما)، بالإضافـة إلى العلاقـات بين المتغيرات، أمرًا بالغ الأهمية. ويمكن أن يكون لتأثير الرسـوم البيانية -والتي تعُد نماذج تخطيطية للنماذج الرياضية - دور في تسـهيل عملية التحديد. كما قد يساعد شـكل أكثر عمومية لرسـم بياني مؤثر، وخريطة معرفية، صانع القرار على تطوير فهم المشـكلة بشكل أفضل، لاسيما المتغيرات وتفاعلاتها.

التنبـؤ (تحليلات توجيهية): وكما ذكرنا سـابقًا، فإن معرفة ما حدث وما هو محتمل حدوثه يُعد شرطًا أساسـيًا هامًا للتحليلات التوضيحية. ويُعتبر هذا النوع من التحليلات التنبؤية ضروريًا

لبنـــاء النـــماذج ومعالجتها نظرًا لأن النتائج المترتبة على تنفيذ القرار عادةً ما تظهر في المســـتقبل. وليس هناك جدوى من إجراء تحليل لمعرفة مدى (الاستجابة) على ما وقع في الماضي لأن القرارات التي ســـيتم اتخاذها في ذلك الوقت لن تُجدي نفعًا في المســـتقبل. وتتطلب التجارة عبر الإنترنت والاتصالات التنبؤ بما قد يحدث وتوافر المعلومات المتاحة من أجل أدائها. وهذه الأنشطة تحدث بسرعة، بعد جمع المعلومات حول المشـــتريات والتي ينبغي تحليلها للوصول إلى التوقعات، حيث ينطوي جزء من التحليل ببساطة على التنبؤ بالطلب؛ ومع ذلك فإن نموذج التنبؤ يمكنه استخدام احتياجات دورة حياة المنتج والمعلومات التي تتعلق بالسوق والمستهلكين لتحليل الوضع بأكمله، مما يؤدي إلى زيادة مبيعات المنتجات والخدمات بشكل مثالي.

وفي الحالة العملية (٦-٢) سوف نقدم وصفًا لمثال فعال لمثل هذه التنبؤات واستخدامها في صنع القرار.

حالة عملية ٦-٢

اســـتخدام تطبيقات ذكاء الأعمال من قِبل شركة إنجرام مايكرو Ingram Micro من أجل اتخاذ قرارات التسعير

تُعد شركـــة إنجرام مايكرو Ingram Micro هي أكبر موزع ثنائي المســـتوى لمنتجات التقنية في العالم. وتقوم الشركة في نظام التوزيع ذي الشقين، بشراء المنتجات من الشركات المصنعة لها ومن ثَم تقوم ببيعها لتجار التجزئة الذين يقومون بدورهم ببيع هذه المنتجات للمســـتخدمين النهائيـــين. فمثلًا، يمكـــن شراء حزمـــة Microsoft Office 365 من شركة إنجرام مايكرو Ingram Micro بدلًا من شرائها مباشرة من مايكروسوفت (Microsoft). وترتبـــط شركة إنجـــرام بالعديد من الشراكات مع شركات مثـــل Best Buy، وBuffalo، و Google، وHoneywell، بالإضافـــة إلى Libratone، وSharper Image. وتقوم الشركة بتسليم منتجاتها إلى ٢٠٠٠٠٠ مزود حلول في جميع أنحاء العالم، وبالتالي فإن الشركة لديها حجـــم كبير من بيانـــات المعاملات. ولتحديد فرص البيع المتقاطع وتحديد الأســـعار التي يتـــم تقديمها لعملاء محددين بالتزامن مع حزم المنتجات، أرادت شركة إنجرام اســـتخدام الإحصـــاءات من هذه البيانات، مما يتطلب إنشـــاء مركز ذكاء الأعـــمال (BIC) من أجل تجميع وتحليل البيانات. وقد واجهت شركة إنجرام العديد من المشاكل في إعداد BIC.

١- واجهت شركة إنجرام مشـــاكل عديدة في عملية التقاط البيانات الخاصة بهم، مثل نقص بيانات الخسائر، وضمان دقة معلومات المستخدم النهائي، وربط الاقتباسات بالأوامر.

٢- أيضًا واجهت شركة إنجرام مشاكل في تنفيذ نظام إدارة علاقات العملاء (CRM) بحيث يكون قادرًا بما يكفي للتعامل مع عملياته في جميع أنحاء العالم.

٣- بالإضافة إلى ما سـبق، فقد واجهت الشركة تصدي لفكرة تسعير الطلب (تحديد السعر وفقًا للإقبال على المنتج).

المنهجية / الحل:

استكشفت شركة إنجرام البريد الإلكتروني واستخدمته في التواصل المباشر مع عملائها (البائعـين) وقدمت لهـم خصومات عـلى شراء التقنيات الداعمـة المتعلقة بالمنتجات المطلوبة. ومن خلال تحليل سـلة السوق المجزأة اسـتطاعوا تحديد هذه الفرص ومن ثَم طوروا تطبيقات ذكاء الأعمال التالية التي سـاعدت في تحديد الأسـعار المُثلى. كما قامت إنجرام بتطوير أداة جديدة لتحسـين الأسعار وتُعرف باسم IMPRIME، وهذه الأداة يمكنها تحديد الأسعار القائمة على البيانات بالإضافة إلى قدرتها على توفير إرشادات التفاوض المسـتندة إلى البيانات، حيث تحدد IMPRIME السـعر الأمثل لكل مستوى من مستويات التسلسل الهرمي للمنتج (مثل: مستوى العميل، ومستوى البائع-العميل، ومستوى شريحة العميل، ومستوى شريحة البائع-العميل). وهي تقوم بفعل ذلك من خلال مراعاة المفاضلة بين إشارة الطلب والتسعير عند ذلك المستوى.

كما قامت الشركة بتطوير برنامج تسويق رقمي يُعرف باسم Intelligence INGRAM، حيـث يسـتخدم هـذا البرنامج التسـجيل الإرشـادي التنبـؤي (PLS)، والـذي يحدد المسـتخدمين النهائيين المُستَهدَفين ببرامج تسـويقية محددة. ويُعد (PLS) هو نظامهم الذي يُسـتخدم لتسـجيل الإرشـادات التنبؤية للشركات التي ليس لها علاقة مباشرة مع المسـتخدمين النهائيين. كما يُستخدم برنامج Intelligence INGRAM لتشغيل برامج الفضاء الأبيض، مما يُشـجع المورِّد على شراء المنتجات المرتبطة ببعضها من خلال تقديم خصومات. فعلى سـبيل المثال، إذا كان المُوَرِّد يشـتري خادمًا من إنجرام، فعندئذٍ تقوم إنجـرام بتقديم خصومات على وحدات تخزين الأقراص نظرًا للترابط بين المنتجين، حيث إنهما مطلوبان للعمل معًا. وبالمثل، فإنه يتم استخدام تقنية Intelligence INGRAM من أجل تطوير الحملات التحفيزيـة (وذلك من خلال تقديم مكافآت نقدية للموردين على تجاوز أهداف الإنفاق الربع سنوية)، وحملات البيع المتبادل (إرسال البريد الإلكتروني للمستخدمين بخصوص المنتجات المرتبطة بالمنتج الذي تم شراؤه مؤخرًا).

النتائج / الفوائد:

وتُسـتخدم منهجية قياس الرفع لقياس الأرباح الناتجة وذلك عن طريق اسـتخدام أداة IMPRIME؛ حيث تقوم هذه المنهجية (منهجية قياس الرفع) بمقارنة الفترات قبل

وبعد تغيير الأســعار كما تقوم بمقارنة مجموعات الاختبــار مقابل مجموعات التحكم. ويتــم إجراء قياس الرفع على متوســط المبيعات اليومية، والهامــش الإجمالي، وهامش الماكينة. وقد أدى استخدام أداة IMPRIME إلى نمو في الإيرادات بلغ ٧٥٧ مليون دولار وزيادة في إجمالي الأرباح بلغت ١٨,٨ مليون دولار.

أسئلة للمناقشة:

١- ما هي أهم التحديات التي واجهتها إنجرام مايكرو Ingram Micro في تطوير BIC؟

٢- اذكر جميع حلول ذكاء الأعمال التي طورتها شركة إنجرام لتحسين أسعار منتجاتها وللتعرف على عملائها.

٣- ما هي الفوائد التي حصلت عليها إنجرام بعد اســتخدام تطبيقات ذكاء الأعمال المطورة حديثًا؟

ما الذي يمكن أن نتعلمه من هذه الحالة العملية؟

تستطيع الشركة أن تبدأ بفهم خطوط منتجاتها، وعملائها، بالإضافة إلى أنماط الشراء الخاصة بهم بشكل أفضل، من خلال بناء BIC لأول مرة. وتُستمَد هذه البصيرة من ما نسميه بالتحليلات الوصفية والتنبؤية. وبالإضافة إلى ذلك، فإن القيمة من هذا تُستَمَد من خلال تحسين الأسعار، وهو أحد مجالات التحليلات التوجيهية.

Sources: Mookherjee, R., Martineau, J., Xu, L., Gullo, M., Zhou, K., Hazlewood, A., Zhang, X., Griarte, F., & Li, N. (2016). End-to-end predictive analytics and optimization in Ingram Micro's two-tier distribution business. Interfaces, 46(1), 4973-; ingrammicrocommerce.com, "CUSTOMERS," https://www.ingrammicrocommerce.com/customers/(accessed July 2016).

فئات النموذج:

يصنف الجدول (٦-١) بعض نماذج القرارات إلى ســبع مجموعات ويسرد العديد من التقنيات الممثلــة لــكل فئة من الفئات؛ حيث يمكن تطبيق كل تقنية إما على نموذج ثابت أو نموذج حركي، والذي يمكن بناؤه في بيئات مفترضة من التأكد، أو عدم التأكد، أو الخطر. ويمكن استخدام أنظمة تحليل القرارات الخاصة والتي يحتوي تصميمها على لغات وإمكانيات نمذجة مدمجة فيها، بغرض تسريــع عملية البناء النموذجية، ويشــمل ذلك جداول البيانات، وأنظمــة التنقيب في البيانات، وأنظمة المعالجة التحليلية عبر الإنترنت (OLAP)، ولغات النمذجة التي تساعد المحلل على بناء

نموذج. وسوف نتناول أحد هذه الأنظمة لاحقًا في هذا الفصل.

إدارة النموذج: وللحفاظ على سلامة النماذج، وبالتالي قابليتها للتطبيق فإنه ينبغي أن تتم إدارة النماذج، مثل البيانات؛ حيث تتم هذه الإدارة بمساعدة أنظمة الإدارة القائمة على النماذج، والتي تشبه أنظمة إدارة قواعد البيانات (DBMS).

النمذجة المستندة إلى المعرفة: تستخدم أنظمة دعم القرار النماذج الكمية في الغالب، في حين تستخدم النظم الخبيرة نماذج نوعية تعتمد على المعرفة في تطبيقاتها. الجدير بالذكر أن بعض المعرفة تكون ضرورية لبناء نماذج قابلة للحل (وبالتالي فإنها تكون قابلة للاستخدام). ولبناء نماذج قائمة على المعرفة، فإنه يمكن استخدام العديد من تقنيات التحليلات التنبؤية، مثل: التصنيف والتجميع.

جدول ٦-١: فئات النماذج

الفئة	العملية والهدف	التقنيات الممثلة
البحث عن الحل الأمثل للمشاكل من خلال بعض البدائل القليلة.	البحث عن أفضل حل من بين عدد قليل من البدائل.	جداول القرارات، وأشجار القرار، وعملية التسلسل الهرمي التحليلي.
البحث عن الحل الأمثل عن طريق خوارزمية.	البحث عن أفضل حل من بين عدد كبير من البدائل، باستخدام عملية التحسين خطوة بخطوة.	نماذج البرمجة الخطية والرياضية الأخرى، ونماذج الشبكات.
البحث عن الحل الأمثل من خلال صيغة تحليلية.	البحث عن أفضل الحلول في خطوة واحدة باستخدام صيغة ما.	بعض نماذج المخزون.
المحاكاة.	البحث عن حل جيد بدرجة كافية أو الحل الأفضل من بين البدائل التي تم فحصها، وذلك من خلال استخدام التجريب.	عدة أنواع من المُحاكاة.
الاستدلال.	العثور على حل جيد بدرجة كافية، عن طريق استخدام القواعد.	برمجة إرشادية، وأنظمة خبيرة.

الفئة	العملية والهدف	التقنيات الممثلة
نماذج تنبؤية.	التنبؤ بالمستقبل بالنسبة لسيناريو معين.	نماذج التنبـؤ، وتحليل ماركوف Markof.
نماذج أخرى.	وضع حـل في حالة "مـاذا - لو"، باستخدام صيغة ما.	النمذجـة الماليـة، وخطـوط الانتظار.

الاتجاهات الحالية في النمذجة: ينطوي أحد الاتجاهات الحديثة في النمذجة على تطوير مكتبات نموذجية ومكتبات تقنية الحل. وبعض هذه الرموز يمكن تشغيلها مباشرةً على خادم الويب الخاص بالمالك مجانًا، والبعض الآخر يمكن تنزيله وتشـغيله على جهاز حاسـب محلي، حيث إن توافر هذه الرموز يعني أن حزم التحسـين والمحاكاة القوية متاحة لصانعي القرار الذين من الممكن أن يقوموا باختبار هذه الأدوات فقط من منظور مشـاكل الفصل الدراسي. فعلى سـبيل المثال، يحتفظ قسـم الرياضيات وعلوم الحاسـب في مختبر أرغـون الوطنيArgonne National (Argonne, Illinois) بخادم NEOS من أجل التحسـين على موقع https://neos-server.org/neos/index.html. كما يمكنـك العثور على روابـط لمواقع أخرى عن طريق النقر على رابط المـوارد في الموقع. Org، الموقع الإلكتروني لمعهد بحوث العمليات وعلوم الإدارة (INFORMS)، حيث يوفر INFORMS معلومات غزيـرة حول النمذجة والحلـول. ويمثل موقع OR/MS Today الموقع الإلكتروني لأحد منشـورات INFORMS، وهو متاح على: http://www.ormstoday.org/ormsmain. كما يتضمن shortly روابط للعديد من فئات برامج النمذجة. وسوف نتناول بعضًا منها قريبًا.

وهنـاك اتجاه واضح نحو تطوير واسـتخدام أدوات وبرامج تسـمى cloud-based من أجل الوصول إلى البرامج وحتى تشغيلها من أجل أداء النمذجة والتحسين والمحاكاة وما إلى ذلك. وهذا يبسط تطبيق العديد من النماذج لمشاكل العالم الحقيقي بطرق عديدة. ومع ذلك، فمن الضروري اكتساب الخبرة جيدًا من خلال تطوير الحلول البسيطة، وذلك من أجل استخدام النماذج وتقنيات الحلول بفاعلية. وغالبًا ما يتم تجاهل هذا الجانب. فالمنظمات التي لديها مُحللين رئيسين والذين يدركون بالفعل كيفية تطبيق النماذج، يطبقونها بفاعلية كبيرة. وهذا يحدث بشـكل ملحوظ في منطقـة إدارة الإيرادات، والتي انتقلت من مقاطعة شركات الطيران والفنادق وتأجير السـيارات إلى التجزئة والتأمين والترفيه والعديد من المجالات الأخرى. كما أن CRM أيضًا تسـتخدم نماذج، غير أنها غالبًا ما تكون شـفافة للمسـتخدم. وتعتبر كمية البيانات وأحجام النماذج كبيرة جدًا مع

النماذج الإدارية، مما يستدعي استخدام مستودعات البيانات لتوفير البيانات وأجهزة الحوسبة المتوازية للحصول على حلول في إطار زمني معقول.

كما أن هناك اتجاهًا مستمرًا نحو جعل نماذج التحليلات واضحة تمامًا لصانع القرار. فعلى سبيل المثال، ينطوي التحليل متعدد الأبعاد (النمذجة) على تحليل البيانات في أبعاد متعددة. وتظهر البيانات بشكل عام في التحليل متعدد الأبعاد (النمذجة) على شكل جدول بيانات، بحيث يكون معظم صناع القرار على دراية بها. ويستخدم الآن العديد من صناع القرار الذين اعتادوا على تشريح وتقطيع مكعبات البيانات، أنظمة OLAP التي تصل إلى مستودعات البيانات. وعلى الرغم من أن هذه الطرق قد تجعل النمذجة مستساغة، إلا أنها أيضًا تخرج العديد من فئات النماذج الهامة والقابلة للتطبيق من الاعتبار، كما أنها تقضي على بعض جوانب تفسير الحلول الهامة والدقيقة. وتتضمن النمذجة ما هو أكثر بكثير من مجرد تحليل البيانات مع خطوط الاتجاه وإقامة علاقات مع الأساليب الإحصائية.

وهناك أيضًا اتجاه لبناء نموذج من نموذج للمساعدة في تحليله بحيث يكون الرسم التخطيطي المؤثر عبارة عن تمثيل بياني لنموذج؛ وهذا هو ما يُقصد به نموذج من نموذج. كما أن بعض حزم برامج الرسم التخطيطي المؤثرة تستطيع إنشاء وحل النموذج الناتج.

أسئلة مراجعة على القسم ٦-٢:

١- اذكر ثلاثة دروس مستفادة من النمذجة.

٢- اذكر مع الوصف المشاكل الرئيسة في النمذجة.

٣- ما هي الأنواع الرئيسة من النماذج المستخدمة في DSS؟

٤- لماذا لا يتم استخدام النماذج في الصناعة بشكل متكرر كما ينبغي أو كما يمكن أن تكون؟

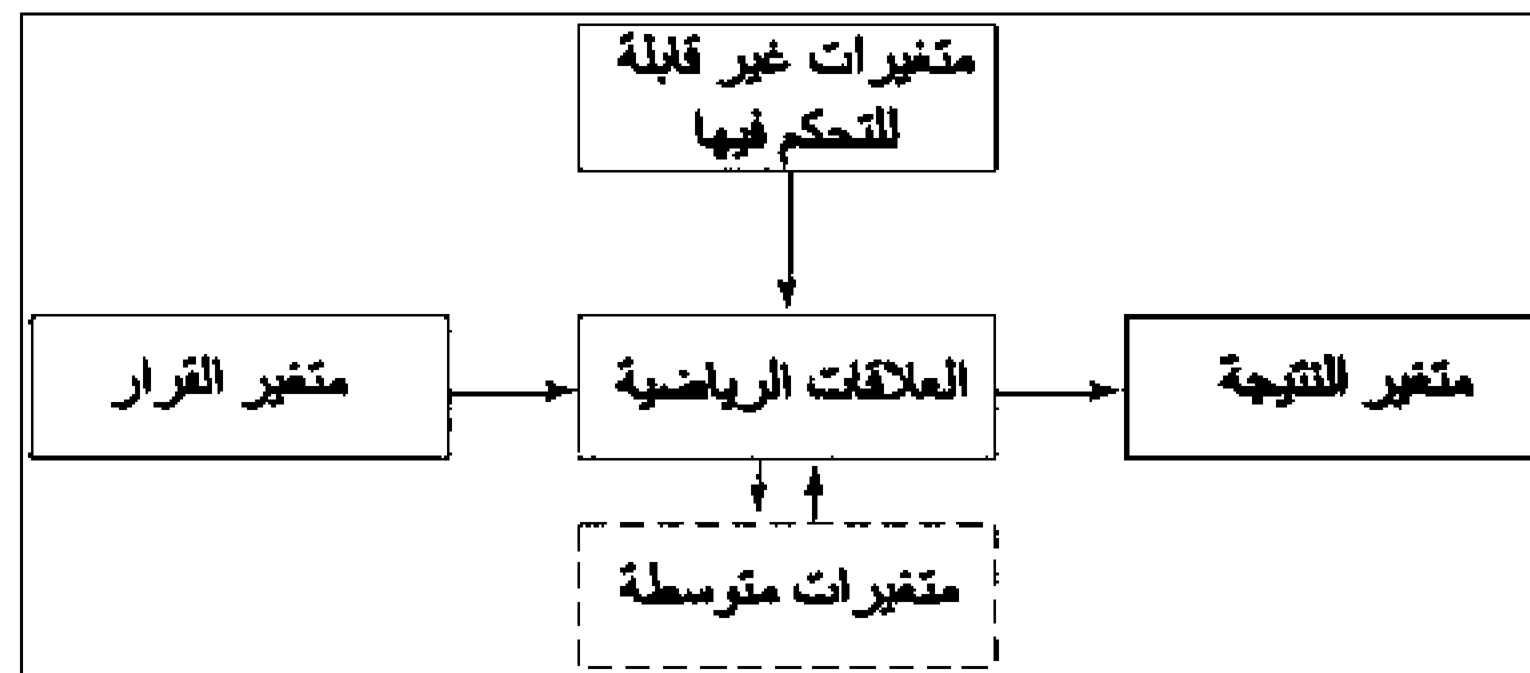
٥- ما هي الاتجاهات الحالية في النمذجة؟

٦-٣ هيكل النماذج الرياضية لدعم القرار:

في الأقسام التالية، سوف نتناول موضوعات النماذج الرياضية التحليلية (كالنماذج الرياضية والمالية والهندسية)، وسوف يتضمن ذلك عناصر وهياكل النماذج.

مكونات النماذج الرياضية لدعم القرار:

تتكون جميع النماذج الكمية عادةً من أربعة عناصر أساسـية (انظر الشكل ٦-١)، وهي: متغيرات النتيجـة (أو الحصيلـة outcome)، ومتغـيرات القرار، والمتغـيرات الغير قابلة للتحكـم (أو المَعْلَمات parameters)، بالإضافـة إلى متغـيرات النتائج المتوسـطة، حيث تقوم العلاقـات الرياضية بربط هذه العنـاصر ببعضهـا البعض. وتكون العلاقات رمزية أو نوعية في النماذج غـير الكمية. وبناءً على القرار المتخذ (أي: قيم متغيرات القرار)، والعوامل التي لا يمكن التحكم فيها من قِبل صانع القرار (في البيئة)، والعلاقات بين المتغيرات، فإنه يتم تحديد نتائج القرارات. وتنطوي عملية النمذجة على تحديد المتغيرات والعلاقات فيما بينها، حيث يُحدد حل النموذج قيم هذه المتغيرات ومتغير (متغيرات) النتيجة.



شكل ٦-١: الهيكل العام لنموذج كمي

متغـيرات النتيجة (الحصيلـة OUTCOME): تعكس متغيرات النتيجـة (الحصيلة outcome) مـدى فاعلية النظام؛ بمعنى أنها تشـير إلى مدى جـودة النظام أو تحقيق هدفـه (أهدافه). وهذه المتغـيرات هـي مخرجات. وقد وردت أمثلة لمتغـيرات النتائج في الجـدول (٦-٢). وتعتبر متغيرات النتائج متغيرات تابعة. كما تُسـتخدم متغيرات النتائج المتوسـطة أحيانًا في النمذجة لتحديد النتائج المتوسطة. وفي حالة المتغير التابع، ينبغي أولًا وقوع حدث آخر قبل وقوع الحدث الموصوف بواسطة المتغير. وتعتمد متغيرات النتائج على حدوث متغيرات القرار والمتغيرات التي لا يمكن التحكم فيها.

متغـيرات القـرار: تصف متغيرات القرار مسـارات بديلة للعمل. ويتحكـم صانع القرار في متغيرات القرار. فعلى سـبيل المثال، بالنسبة لمشـكلة الاستثمار، فإن مبلغ الاستثمار في السندات هو متغير القرار. وفي مشكلة الجدولة، فإن متغيرات القرار تكون أشخاصًا وأوقاتًا وجداول. وقد وردت أمثلة أخرى في الجدول (٦-٢).

المتغـيرات الغير قابلة للتحكم، أو المَعْلَـمَات parameters: هناك عوامل تؤثر على متغيرات النتائج في أي موقف لاتخاذ القرار، غير أن هذه العوامل خارج نطاق سيطرة صانع القرار. وهذه

العوامـل إما أن تكـون ثابتة، وفي هذه الحالة تسـمى متغيرات غير قابلة للتحكـم، أو مَعْلَمَات parameters، وإما أن تكون قابلة للتغير، وفي هذه الحالة تسمى المتغيرات. ومن أمثلة العوامل سـعر الفائدة الأسـاسي ورمز المباني في المدينة واللوائح الضريبية وتكاليف المرافق. ومعظم هذه العوامـل لا يمكن التحكـم فيها نظرًا لكونها موجودة ومحددة من قِبـل عناصر بيئة النظام التي يعمل فيها صانع القرار. كما أن بعض هذه المتغيرات يُقَيِّد صانع القرار، وبالتالي فإنها تُشَـكل ما يُسمى بالقيود المفروضة على المشكلة.

متغيرات النتائج المتوسـطة: تعكس متغيرات النتائج المتوسطة الحصيلة المتوسطة في النماذج الرياضيـة. فمثلًا، عند تحديد جدولة الآلة، فإن التلف يكون هو متوسـط متغير الناتج، ويكون الربح الإجمالي هو متغير النتيجة (أي إن التلف هو أحد العوامل المحددة لإجمالي الربح). ومثال آخر: هو رواتب الموظفين، ويُشـكل هذا متغير قرار الإدارة؛ إذ يحدد مدى ارتياح الموظفين (أي: الحصيلة المتوسطة)، والذي يحدد بدوره مستوى الإنتاجية (أي: النتيجة النهائية).

جدول ٦-٢: أمثلة لعناصر النماذج

المنطقة	متغيرات القرار	متغيرات النتيجة	متغيرات ومَعْلَمَات غير قابلة للتحكم
الاستثمار المالي	بدائل الاستثمار ومبالغه	مجموع الربح، والخطر معدل العائد على الاستثمار (ROL) ربحية السهم مستوى السيولة	معدل التضخم معدل الفائدة المنافسة
التسويق	ميزانية الإعلان مكان الإعلان	الحصة التسويقية رضا العملاء	دخْل العميل إجراءات المنافس
التصنيع	نوعية المنتج وكميته مستويات المخزون برامج التعويض	التكلفة الإجمالية مستوى الجودة ارتياح الموظف	سعة الماكينة التقنية أسعار المواد
المحاسبة	استخدام أجهزة الحاسب جدول التدقيق	تكلفة معالجة البيانات نسبة الخطأ	تقنية الحاسب معدلات الضرائب المتطلبات القانونية

المنطقة	متغيرات القرار	متغيرات النتيجة	متغيرات ومَعْلَمَات غير قابلة للتحكم
وسائل النقل	جدول الشحنات استخدام البطاقات الذكية	مجموع تكلفة النقل قيمة وقت الفراغ Payment float time	مسافة التسليم القوانين
الخدمات	مستويات التوظيف	ارتياح العملاء	الطلب على الخدمات

هيكل النماذج الرياضية لدعم القرار:

ترتبط عناصر النموذج الكمي بواسطة التعبيرات الرياضية (الجبرية) - المعادلات أو التباين.

وفيما يلي نموذج مالي بسيط جدًا $P = R - C$.

حيث P = الربح، وR = الإيرادات، وC = التكلفة. وتصف هذه المعادلة العلاقة بين المتغيرات. وهناك نموذج مالي آخر معروف هو نموذج التدفق النقدي البسيط للقيمة الحالية، حيث P = القيمة الحالية، وF = دفعة واحدة مستقبلية بالدولار، وi = سعر الفائدة (النسبة المئوية)، وn= عدد السنوات. ويمكن مع هذا النموذج تحديد القيمة الحالية لدفع مبلغ ١٠٠,٠٠٠$ يتم سداده بعد ٥ سنوات من اليوم، بمعدل فائدة ١٠٪ (٠,١)، على النحو التالي: P = ١٠٠,٠٠٠ / (١ + ٠,١) ٥ = ٦٢,٠٩٢.

وفيما يلي من أقسام سوف نقدم نماذج رياضية أكثر تشويقًا وتعقيدًا.

أسئلة مراجعة على القسم ٦-٣:

١- ما هو متغير القرار؟

٢- اذكر وناقش بإيجاز العناصر الرئيسة للنموذج الكمي.

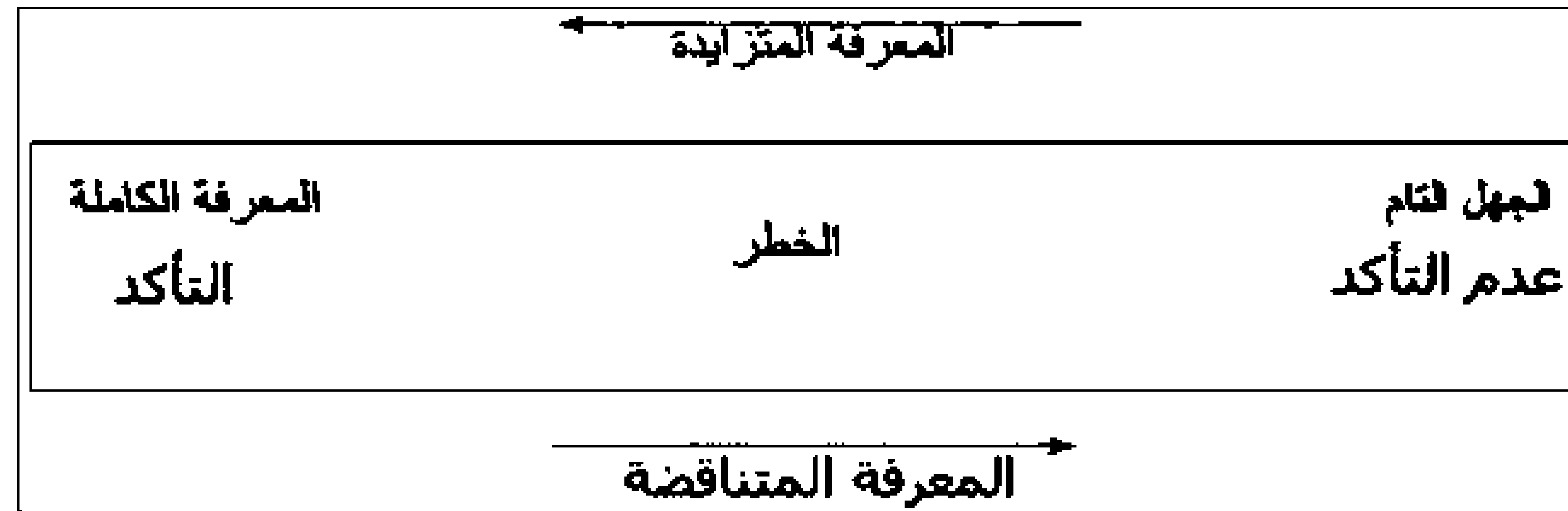
٣- اشرح دور متغيرات النتائج المتوسطة.

٦-٤ التأكد، وعدم التأكد، والخطر:

تنطوي عملية اتخاذ القرار على تقييم ومقارنة البدائل. ومن الضروري خلال هذه العملية، التنبؤ بالنتائج المستقبلية لكل بديل مقترح. وغالبًا ما يتم تصنيف حالات القرار على أساس ما يعرفه (أو ما يعتقده) صانع القرار بشأن النتائج المتوقعة. وعادةً ما نقوم بتصنيف هذه المعرفة إلى ثلاث فئات (انظر الشكل ٦-٢)، بدءًا من المعرفة الكاملة إلى الجهل الكامل، وهي:

- التأكد. - عدم التأكد. - الخطر.

وقــد تحــدث أي حالة من هذه الحــالات عندما نقــوم بتطوير النماذج، ولــكل حالة أنواع مختلفة من النماذج تكون مناسبة لها. وفيما يلي سوف نناقش كلًا من التعريفات الأساسية لهذه المصطلحات وبعض قضايا النمذجة الهامة لكل حالة.



شكل ٦-٢: مناطق اتخاذ القرار

اتخاذ القرار في ظل التأكد:

مــن المفترض أن المعرفة الكاملة تكون متاحة بالنســبة لصانع القرار عند اتخاذ القرار في ظل التأكد، بحيث يكون صانع القرار على علم تام بنتيجة كل مســار من العمل (كما لو كان في بيئة حتميــة). وقد لا يكــون صحيحًا أن معرفة النتائج تكون بنســبة ١٠٠٪، كما لا يلزم إجراء تقييم حقيقي لكل النتائج، غير أنه غالبًا ما يقوم هذا الافتراض بتبســيط النموذج ويجعله قابلًا للتتبع، حيث يُنظر إلى صانع القرار على اعتبار أنه متنبئ مثالي للمستقبل نظرًا لأنه من المفترض أن هناك نتيجة واحدة فقط لكل بديل. فعلى ســبيل المثال، بديل الاســتثمار في سندات الخزانة الأمريكية هــو الذي تتوفر به معلومات كاملة حول عائد الاســتثمار المســتقبلي إذا تــم الاحتفاظ به حتى الاســتحقاق. وغالبًا ما تحدث حالة اتخاذ القرار في ظل التأكد، مع مشــاكل هيكلية وآفاق زمنية قصيرة (تصل إلى ســنة واحدة). وتُعد النماذج القائمة على التأكد سهلة نسبيًا في تطويرها وحلها، كما يمكنها تقديم الحلول المُثلى. ويتم إنشاء العديد من النماذج المالية في ظل تأكد مفترض، على الرغم من أن السوق لا يقبل إلا بدرجة تأكد ١٠٠٪.

اتخاذ القرار في ظل عدم التأكد:

في حالــة اتخاذ القرار في ظل عدم التأكد، فإن صانع القرار يأخذ في الاعتبار الحالات التي تكون فيها عدة نتائج ممكنة لكل مسار عمل. وهذه الحالة على النقيض من وضع الخطر، حيث إن صانع

القــرار في هــذه الحالة لا يعرف، أو لا يمكنه تقدير احتمالية حــدوث النتائج الممكنة. ويُعد اتخاذ القرار في ظل عدم التأكد أكثر صعوبة من اتخاذه في ظل التأكد بسبب عدم توافر معلومات كافية. وينطوي وضع النماذج لمثل هذه الحالات على تقييم موقف صانع القرار (أو المنظمة) تجاه الخطر.

ويحاول المديرون تجنب حالة عدم التأكد قدر الإمكان، إلى الدرجة التي تجعلهم يستبعدونه حتــى من مجرد الافتراض. فهم يحاولون الحصول على مزيد من المعلومات بحيث يمكن التعامل مع المشــكلة في ظل التأكد (لأنه قد يكون «شبه» مؤكد) أو تحت خطر محسوب (أي: مفترض). وفي حالة عدم توفر المزيد من المعلومات، ينبغي معالجة المشــكلة تبعًا لحالة عدم التأكد، والتي قد تكون أقل تحديدًا من الفئات الأخرى.

اتخاذ القرار في ظل الخطر (تحليل الخطر):

يُقصد بالقرار الذي يُتخذ في ظل الخطر (المعروف أيضًا بحالة اتخاذ القرار الاحتمالي أو العشوائي) بأنــه القرار الذي يجب أن يأخذ فيه صانع القــرار بعين الاعتبار عدة نتائج ممكنة لكل بديل، ولكل منهــا احتمال حــدوث معين. ومن المفترض أن تكون احتمالات حــدوث النتائج المحددة على المدى الطويل معروفة أو يمكن تقديرها. ومن خلال هذه الافتراضات، يســتطيع صانع القرار تقييم درجة الخطــر المرتبط بكل بديل (يســمى الخطر المحســوب). ومن الجدير بالذكــر أن معظم القرارات التجارية الرئيســة يتم اتخاذها في ظل الخطر المفترض. أما تحليل الخطر (أي: الخطر المحســوب) فهي طريقة صنع القرار التي تحلل الخطر (بناءً على الاحتمالات المعروفة المفترضة) المرتبطة بالبدائل المختلفة. ويمكن إجراء تحليل الخطر عن طريق حســاب القيمة المتوقعة لكل بديل واختيار البديل ذي القيمة المتوقعة الأفضل. وتوضح الحالة العملية (٦-٣) تطبيقًا واحدًا لتقليل عدم التأكد.

حالة عملية ٦-٣

اســتخدام الخطــوط الجوية الأمريكية نمذجة التكلفة اللازمــة لتقييم عدم التأكد من عطاءات طرق الشحن

تُعــد شركة الخطوط الجوية الأمريكية American Airlines, Inc (AA)إحدى أكبر شركات الطيران في العالم. وأما الأعمال الأساســية للشركة فتتمثل في نقل الركاب، بالإضافة إلى أن لديها وظائف فرعية حيوية أخرى تتضمن الشــحن الكامل للشاحنات (FTL) من معدات الصيانة وشــحنات بنود خدمة المســافرين والتي يمكنها إضافة ما يزيد عن مليار دولار في المخزون في أي وقت. وتتلقى AA عروض عديدة من الموردين استجابةً لطلبات

عروض الأسعار requests for quotes (RFQs) من أجل قوائم الجرد. وقد يزيد إجمالي طلبات عروض الأسعار RFQs لشركة AA عن ٥٠٠ طلب عرض أسعار في أي عام. ونتيجة للعدد الكبير من عروض التسـعير وعملية تقديم العطاءات المعقدة الناتجة، فإن عروض أسعار الإعلانات تختلف بشكل كبير. وفي بعض الأحيان، فإنه يمكن لعقد واحد الانحراف بنحو ٢٠٠٪. ونتيجة لهذه العملية المعقدة، فإنه من الشائع قيام الموردين إما بدفع مبالغ زائدة أو عدم دفعها من أجل الحصول على خدماتهم. وتحقيقًا لهذه الغاية، فقد أرادت شركة الخطوط الأمريكية الجوية بناء نموذج شراء من شأنه تبسيط وتقييم عروض الأسعار من الموردين من أجل اختيار عروض أسعار عادلة لكلا الطرفين ومورديهم.

المنهجية / الحل:

وقد تم اتخاذ ثلاث خطوات، لتحديد التكلفة العادلة لمنتجات وخدمات الموردين، وهي:

١- استقصاء المصادر الأساسية (كالمقابلات) والثانوية (كالإنترنت) بحثًا عن الحالة الأساسية وبيانات المجال والتي من شأنها الإخبار عن متغيرات التكلفة التي تؤثر على عرض FTL.

٢- اختيار متغيرات التكلفة بحيث تكون حصرية تبادلية وجماعية شاملة.

٣- استخدام برنامج تحليل قرار DPL لتوضيح عدم التأكد.

وعلاوةً على ذلك، فقد اسـتخدمت Swanson Megill التقريب الموسـع لنمذجة توزيع الاحتمالات للمتغيرات الأكثر حساسـية من حيث التكلفة المسـتخدمة. وقد تم ذلك لحساب التقلبات العالية في عروض الأسعار في النموذج الأولي.

النتائج / الفوائد:

تم إجراء اختبار تجريبي على طلبات عروض الأسـعار التي جذبت عروض من ست ناقـلات FTL. وقد كان من بين العروض السـت المقدمة خمسـة عروض ضمن ثلاثة انحرافات معيارية عن المتوسط، في حين استُبْعِدَ السادس. وقد استخدمت AA في وقت لاحـق نموذج FTL للتكلفـة اللازمة لأكثر من ٢٠ طلب سـعر RFQs من أجل تحديد التكلفة العادلة والدقيقة للسـلع والخدمات. ومن المتوقع أن يسـاعد هذا النموذج في الحد من الأخطار التي تتعلق إما بالدفع الزائد أو دفع تعويضات للموردين.

أسئلة للمناقشة:

١- إلى جانب الحد من خطر الدفع الزائد أو دفع تعويضات للموردين، ما هي بعض المميزات الأخرى التي تستطيع شركة AA أن تستمدها من «نموذج ما ينبغي أن يكون»؟

٢- هــل يمكنك التفكير في المجالات الأخرى التي يمكن اســتخدام هذا النموذج فيها إلى جانب مجال النقل الجوي؟

٣- ناقش الطرق الأخرى الممكنة التي تستطيع من خلالها شركة AA أن تحل مشكلة دفع مبالغ زائدة ومقدار المدفوعات.

Source: Adapted from Bailey, M. J., Snapp, J., Yetur, S., Stonebraker, S., Edwards, S. A., Davis, A., & Cox, R. (2011). Practice summaries: American Airlines uses should-cost modeling to assess the uncertainty of bids for its full-truckload shipment routes. Interfaces, 41(2), 194196-.

أسئلة مراجعة على القسم ٦-٤:

١- وضح معنى القيام بعملية اتخاذ القرار في ظل التأكد المفترض، والخطر، وعدم التأكد.

٢- كيف يمكن التعامل مع مشاكل صنع القرار في ظل التأكد المفترض؟

٣- كيف يمكن معالجة مشاكل صنع القرار في ظل عدم التأكد المفترض؟

٤- كيف يمكن التعامل مع مشاكل صنع القرار في ظل الخطر المفترض؟

٦-٥ نمذجة القرار باستخدام جداول البيانات:

يمكن تطوير النماذج وتنفيذها في مجموعة متنوعة من لغات البرمجة والأنظمة. ونحن نركز في المقــام الأول على جداول البيانات (مــع إضافاتهم)، ونماذج اللغات، بالإضافة إلى أدوات تحليل البيانــات الواضحة. وسرعان ما تم الاعتراف بحزم جــداول البيانات باعتبارها برنامج تنفيذ يمكن اســتخدامه بســهولة من أجل تطوير مجموعة واسعة من التطبيقات في مجال الأعمال والهندسة والرياضيات والعلوم. وتشــتمل جداول البيانات على الإحصائيات والتوقعات الشــاملة، بالإضافة إلى التصميمات الأخرى، وقدرات إدارة قواعد البيانات، والمهام، والروتين. وقد تم تطوير الوظائف الإضافيــة مع تطور حزم جــداول البيانات، وذلك من أجل هيكلة وحل فئات معينة من النماذج. ومــن بين الحزم المضافة، تــم تطوير العديد منها من أجل تطوير DSS. وتتضمن هذه الوظائف الإضافية المرتبطة بأنظمة دعم القرار كلاً من:

- Solver (Frontline Systems Inc., solver.com).

- What'sBest! (إصــدار Lindo، من أنظمة لينــدو Lindo Systems Inc., lindo.com;) من أجل الأداء الخطي والغير خطي الأمثل.

- Braincel (Jurik Research Software, Inc., jurikres.com).
- الأدوات العصبية Neural Tools (Palisade Corp., palisade.com).
- الشبكات العصبية الاصطناعية artificial neural networks.
- المُطَوِّر Evolver (Palisade Corp) للخوارزميات الجينية.
- الخطر RISK@ (Palisade Corp) لأداء دراسات المحاكاة.

وتتوفــر الوظائف الإضافية القابلة للمقارنة مجانًـا أو بتكلفة منخفضة جدًا. (يمكنك إجراء بحث على الويب حتى تتمكن من الوصول إليها؛ فهي بمثابة إضافة جديدة منتظمة إلى السوق). ونظرًا لأن جـدول البيانات يضم العديد من الوظائف الماليـة والإحصائية والرياضية وغيرها من الوظائف القوية، فهو يُعد أداة النمذجة الأوسـع انتشـارًا للمستخدم النهائي. وتستطيع جداول البيانات تنفيذ مهام الحلول النموذجية مثل البرمجة الخطية وتحليل الانحدار. وقد تطور جدول البيانـات إلى أداة مهمة للتحليل والتخطيط والنمذجة (انظر Farasyn, Perkoz, & Van de Velde, 2008; Hurley & Balez, 2008; Ovchinnikov & Milner, 2008). وتقـدم الحالتان العمليتان (٦-٤) و(٦-٥) تطبيقات مثيرة للاهتمام للنماذج المستندة إلى جداول البيانات في بيئة غير ربحية.

حالة عملية ٦-٤

اسـتخدام بنسيلفينيا أدوبشن إكسشـانج (Pennsylvania Adoption Exchange) لنموذج جدول البيانات من أجل تحسين تطابق الأطفال مع الأسر

تم تأسـيس Pennsylvania Adoption Exchange (PAE) في عام ١٩٧٩ من قِبل ولاية بنسلفانيا لمساعدة المقاطعات والوكالات غير الهادفة للربح في العثور على عائلات محتملة للأطفال اليتامى الذين لم يقم أحد بتبنِّيهم بسبب العمر أو الاحتياجات الخاصة. وتحتفظ PAE بسجلات مفصلة عن الأطفال وتفضيلات الأسر التي من المحتمل أن تقوم بتبنِّيهم. كما تبحث عن عائلات للأطفال في جميع المقاطعات الـ٦٧ في ولاية بنسلفانيا.

وتُعد شـبكة بنسلفانيا للتبنِّي والاستدامة هي المسؤولة على مستوى الولاية عن إيجاد منازل دائمة للأيتام. فإذا فشلت الشبكة بعد عدة محاولات في وضع طفل مع عائلة، فإنها حينئذٍ تحصل على مسـاعدة من PAE. وتسـتخدم PAE أداة تقييم آلية لتنسـيق وضع الأطفال مع العائلات المناسـبة لهـم، حيث تقدم هذه الأداة توصيـات للتوفيق من خلال حسـاب النتيجة بـين ٠ و١٠٠٪ للطفل على ٧٨ زوجًا من قيم سـمات الطفل والتفضيلات

العائليـــة. وقد كافحت PAE لعدة ســنوات مــن أجل إعطاء توصيــات توفيق التبنِّي بين الأطفال والأسر التي تناســبهم للأخصائيين المســؤولين عن حالات الأطفال. غير أنهم وجدوا صعوبــة في إدارة قاعــدة بيانات واســعة من الأطفــال التي تم جمعها بمــرور الوقت على مســتوى جميع المقاطعات الـ٦٧. وقد قدمت خوارزمية البحث الرئيسة توصيات للتوفيق أثبتــت أنها غير مثمرة للأخصائيين المســؤولين عن الحالات. ونتيجــة لذلك؛ فقد زاد عدد الأطفال الذين لم يتم تبنِّيهم زيادة هائلة، وأصبحت الحاجة ملحة لإيجاد أسر لهؤلاء الأيتام.

المنهجية / الحل:

بدأت PAE بجمع المعلومات عن كل من الأيتام والعائلات عن طريق الاستطلاعات عبر الإنترنت والتي تشــمل مجموعة جديدة من الأسئلة؛ حيث تعمل هذه الأسئلة على جمع المعلومات الخاصة بهوايات الطفل وتفضيلات الأخصائيين المســؤولين عن الأطفال للعائــلات، وفئة الأطفال العمرية المفضلة من قبل العائلات. وقد أنشــأ PAE والأطباء الأخصائيون أداة لتوفيق جداول البيانات، وقد اشتملت هذه الأداة على مميزات إضافية إذا ما قورنت بالأداة الإضافية التي استُخدِمت سابقًا. ويستطيع الأخصائيون المسؤولون عــن الحالات في هــذا النموذج تحديد أهمية الســمات من أجل اختيــار عائلة لطفل. فمثــلًا، إذا كانت إحدى العائلات لديها مجموعــة محدودة من التفضيلات التي تتعلق بنــوع الطفل، وعمــره، وعرقه، فإن هــذه العوامل يمكن أن تصبــح ذات أهمية أعلى. وبالإضافة إلى ذلك، فإن الأخصائيين المســؤولين عن الحالات يمكنهم أن يعطوا الأفضلية حــول المقاطعة التي تقيم بهــا الأسرة، نظرًا لأهمية عامل العلاقة الاجتماعية بالنســبة للطفل. وتســتطيع لجنة التوفيق باســتخدام هذه الأداة مقارنة الطفل والأسرة في كل سمة، مما يؤدي إلى جعل قرار التوفيق بين كل من الأسرة والطفل أكثر دقة.

النتائج / الفوائد:

منذ أن بدأت PAE باســتخدام نموذج جدول البيانات الجديد من أجل التوفيق بين عائلة وأحد الأطفال، استطاعوا اتخاذ قرارات خاصة بالتوفيق أفضل، مما أدى إلى ارتفاع نسبة الأطفال الذين يحصلون على منزل دائم.

وتمثل هذه الحالة القصيرة مثالًا واحدًا من العديد من الأمثلة التي تستخدم جداول البيانــات كأداة لدعم القرار، حيث يتم إنشــاء نظام مطابقــة أفضل بين رغبات العائلة التي ترغب في التبنِّي وسمات الطفل، من خلال إنشاء نظام تسجيل بسيط لرغبة العائلة وسمة الطفل بحيث يتم الإبلاغ عن عدد أقل من حالات الرفض لأيٍ من الجانبين.

أسئلة للمناقشة:

١- ما هي التحديات التي واجهتها PAE أثناء اتخاذ قرارات توفيق حالات التبنِّي؟

٢- ما هي مميزات أداة جدول البيانات الجديدة التي ساعدت PAE في حل المشاكل التي تتعلق بالتوفيق بين عائلة وطفل؟

Source: Adapted from Slaugh, V. W., Akan, M., Kesten, O., & Unver, M. U. (2016). The Pennsylvania Adoption Exchange improves its matching process. Interfaces, 46(2), 133154-.

حالة عملية ٦-٥

استخدام مؤسسة مترو ميلز أون ويلز تريجر فالي (Metro Meals on Wheels Treasure Valley) لبرنامج Excel من أجل البحث عن خطوط السير المُثلى للتسليم

Meals on Wheels Association of America (وتسمى الآن Meals on Wheels America) وهي منظمة غير ربحية تقوم بتقديم ما يقرب من مليون وجبة إلى منازل كبار السن المحتاجين عبر الولايات المتحدة. ويُعد Metro Meals on Wheels Treasure Valley فرعًا محليًّا لـ Meals on Wheels America يعمل في إيداهو (Idaho)، حيث يحتوي هذا الفرع على فريق من السائقين المتطوعين الذين يستخدمون سياراتهم الشخصية كل يوم من أجل تقديم الوجبات إلى ٨٠٠ عميل على طول ٢١ مسارًا ويغطون مساحة تبلغ ٢,٧٤٥ كيلومتر مربع.

وقد واجهت منظمة Meals on Wheels Treasure Valley العديد من المشاكل. أما المشكلة الأولى، فقد تمثلت في تطلعهم إلى تقليل وقت التسليم حيث كان الطعام المطبوخ حساسًا لدرجة الحرارة ومن الممكن أن يفسد بسهولة. وقد كانوا يرغبون في توصيل الطعام المطبوخ في غضون ٩٠ دقيقة من بعد مغادرة السائق من أجل توصيل الطعام. وأما المشكلة الثانية، فقد تمثلت في استغراق عملية الجدولة وقتًا طويلًا جدًا. وقد قضى اثنان من الموظفين الكثير من وقتهم في تطوير مسارات مجدولة من أجل التوصيل. وقد قدم منسق خطوط السير بتحديد عدد مرات التوقف وفقًا لعدد متلقي الوجبات ليوم معين. وبعد تحديد نقاط التوقف، قام المنسق بعمل سلسلة من التوقفات التي قللت من وقت

سـفر المتطوعين. وبعد ذلك يتم إدخال جدول خط السـير في أداة عبر الإنترنت لتحويل تعليمات القيادة للسائقين. وقد كانت عملية تحديد خطوط السير يدويًا تسـتغرق الكثير من الوقت الإضافي. وتحتاج MMW إلى أداة تمكنها من تحسين نظـام التوصيـل الخاص بهم فضلًا عن قدرتها على تقديـم حلول توجيه لكل من اتجاهات الذهاب والعودة لتوصيل وجبات الطعام. ويمكن لأولئك الذين يقودون بانتظام أن يقوموا بتوصيل الطعام الدافئ أو البارد في اليوم التالي. أما الأشـخاص الآخرون الذين يقومون بقيادة السـيارات فقط في بعض الأحيان، فإنهم يحتاجون إلى العودة إلى المطبخ أحيانًا لتوصيل الأطعمة الدافئة / المبردة.

المنهجية / الحل:

تـم تطوير أداة تعتمد على جدول البيانات، من أجل حل مشـكلة خط السـير. حيث تحتوي هذه الأداة على واجهة لإدخال المعلومات الخاصة بالمسـتلم بسـهولة مثل اسـمه / اسـمها ومتطلبات الوجبة، وعنوان التسـليم، حيث ينبغي ملء هذه المعلومـات في جـدول البيانات لكل نقطة توقف في خط السـير، ثـم بعد ذلك تم اسـتخدام برنامج Excel's Visual Basic لوظيفـة التطبيقات من أجل الوصول إلى واجهة برمجة تطبيقات خرائط الشبكة للمطورين (API) والتي تسمى MapQuest. وقد تم اسـتخدام هذه الواجهة الخاصة ببرمجة التطبيقات API لإنشـاء مصفوفة سفر تقوم بحساب الوقت والمسافة اللازمة لتوصيل الوجبة، حيث تعطي هذه الأداة معلومات عن الوقت والمسافة لـ٥٠٠٠ زوج من المواقع يوميًا بدون أي تكلفة.

وعند بدء التشـغيل يقـوم MapQuest API أولًا بالتأكد من صحة العناوين المدخلة لمتلقي الوجبات. ثم يسـتخدم البرنامج واجهة برمجة التطبيقات (API) للحصول على مسـافة القيادة، والوقت المتوقع للقيادة، وتغيير التعليمات للقيادة بين جميع نقاط التوقف في الطريق. وبعد ذلك، تسـتطيع الأداة الوصول إلى ٣٠ محطة خلال فترة زمنية ممكنة.

النتائج / الفوائد:

أدى اسـتخدام هذه الأداة، إلى انخفاض إجمالي مسافة القيادة السنوية بمقدار ١٠,٠٠٠ ميل، بينما قل وقت السفر بمقدار ٥٣٠ ساعة. وقد وفرت وجبات Metro Meals on Wheels Treasure Valley ٥,٨٠٠$ في عام ٢٠١٥، اعتمادًا على معدل

توفير يقدر بـ ٠,٥٨$ (لكل سـيارة سيدان متوسـطة الحجم). كما استطاعت هذه الأداة تقليل الوقت المستغرق في تخطيط خط السير لتوصيل الوجبات. كما تضمنت المزايا الأخرى زيادة رضا المتطوعين بالإضافة إلى زيادة فرص الاحتفاظ بهم.

أسئلة للمناقشة:

١- ما هـي التحديـات المتعلقـة بتوصيـل الوجبـة التـي واجهـت Metro Meals on Wheels Treasure Valley related قبـل اعتماد الأداة المستندة إلى جدول البيانات؟

٢- وضح تصميم نموذج يستند إلى جدول البيانات.

٣- ما هي الفوائد غير الملموسـة التي حصلت عليها Metro Meals on Wheels بسبب استخدام نموذج يستند إلى Excel؟

Source: Adapted from Manikas, A. S., Kroes, J. R., & Gattiker, T. F. (2016). Metro Meals on Wheels Treasure Valley employs a low- cost routing tool to improve deliveries. Interfaces, 46(2), 154167-.

تنطـوي المميـزات المهمة الأخرى لجـدول البيانات على تحليلات "ماذا - لـو"، والبحث عن الهدف، بالإضافة إلى إدارة البيانات، وقابلية البرمجة (أي وحدات الماكرو). ويمكن بسـهولة تغيير قيمة الخلية والاطلاع فورًا على النتيجة، باسـتخدام جـدول البيانات. كما يتم تنفيذ البحث عن الهدف من خلال الإشـارة إلى خلية الهدف والقيمة المطلوبة والخلية المتغيرة. ويمكن تنفيذ إدارة قاعدة بيانات واسـعة مع مجموعات بيانات صغيرة، أو أجزاء من قاعدة بيانات يمكن اسـتيرادها للتحليل (والذي هو في الأسـاس كيفية عمل OLAP مع مكعبات البيانات متعددة الأبعاد. وفي الواقع، فإن معظم أنظمة OLAP لديها الشكل والمظهر من برامج جداول البيانات المتقدمة بعد تنزيل البيانات)، وتقوم قوالب ووحدات الماكرو والأدوات الأخرى بتعزيز إنتاجية بناء DSS.

ونظرًا لأن حزم جداول البيانات تقرأ وتكتب بنى ملفات شـائعة وتتفاعل بسهولة مع قواعد البيانات والأدوات الأخرى؛ فإن معظمها يوفر تكاملًا سلسًا إلى حدٍ ما. وتُعد مايكروسوفت إكسل هي حزمة جدول البيانات الأكثر انتشـارًا. وفي الشـكل (٦-٣)، نقوم بعرض نموذج حساب قرض بسيط حيث تصف المربعات في جدول البيانات محتويات الخلايا، والتي تحتوي على صيغ؛ حيث يظهـر تغيـير في معدل الفائدة في الخلية E٧ على الفور في الدفع الشـهري في الخلية E١٣. ومن الممكن ملاحظة النتائج وتحليلها على الفور. فإذا كنا بحاجة إلى دفعة شـهرية محددة، فيمكننا استخدام البحث عن الهدف بغرض تحديد سعر فائدة مناسب أو مبلغ قرض.

الجدير بالذكر أن جدول الحســابات الموضح في الشــكل (٦-٣) ثابت. وعلى الرغم من تأثير المشــكلة على المقترض بمرور الوقت، إلا أن النموذج يشــير إلى أداء شهر واحد، والذي يتم تكراره. وعلى النقيض، فإن النموذج الحركي يمثل الســلوك بمرور الوقت. كما تشــير حسابات القروض في جدول البيانات الموضحة في الشكل (٦-٤) إلى تأثير الدفع المسبق على المبلغ الأصلي بمرور الوقت. وباســتخدام مولدات رقمية عشوائية مدمجة لتطوير نماذج المحاكاة، يمكن دمج تحليل الخطر في جداول البيانات (انظر الفصل التالي).

Simple Loan Calculation Model in Excel

Loan Amount $150,000
Interest Rate 8.00%
Number of Years 30
Number of Months 360 =E8*12
Interest Rate/Month 0.67% =E7/12
Monthly Loan Payment $1,100.65 =PMT (E11, E10, E6, 0)

Excel Spreadsheet Static Model Example of a Simple Loan

شكل ٦-٣: مثال نموذج لجداول Excel الإحصائية لحساب قرض بسيط بمدفوعات شهرية

Dynamic Loan Calculation Model with Prepayment in Excel

Loan Amount $150,000
Interest Rate 8.00%
Number of Years 30
Number of Months 360
Interest Rate/Month 0.67%
Monthly Loan Payment $1,100.65

Excel Spreadsheet: Dynamic Model Example of a Simple Loan

Month	Normal Payment	Prepay Amount	Total Payment	Principle Owed
0				$150,000
1	$1,100.65	$100.00	$1,200.65	$149,799
2	$1,100.65	$100.00	$1,200.65	$149,597
3	$1,100.65	$100.00	$1,200.65	$149,394
4	$1,100.65	$100.00	$1,200.65	$149,189
5	$1,100.65	$100.00	$1,200.65	$148,983

شكل ٦-٤: مثال نموذج ديناميكي لجداول بيانات Excel لحساب قرض بسيط بمدفوعات شهرية وتأثيرات الدفع المسبق

ويتم الإبلاغ عن تطبيقات جداول البيانات للنماذج بانتظام. وفي القسم التالي سوف نتعلم كيفية استخدام نموذج تحسين يعتمد على جدول البيانات.

أسئلة مراجعة على القسم ٦-٥:

١- ما هو جدول البيانات؟

٢- ماذا أضاف جدول البيانات؟ وكيف تساعد هذه الإضافة في إنشاء واستخدام DSS؟

٣- اشرح لماذا يؤدي جدول البيانات إلى تطوير نظام DSS.

٦-٦ تحسين البرمجة الرياضية:

إن البرمجة الرياضية هي مجموعة من الأدوات التي صُممت من أجل المساعدة في حل المشاكل الإدارية والتي ينبغي أن يقوم فيها صانع القرار بتخصيص موارد نادرة بين الأنشطة المتنافسة من أجل تحسين هدف قابل للقياس. فعلى سبيل المثال، يعتبر توزيع وقت الماكينة (المورد) بين المنتجات المختلفة (الأنشطة) مشكلة تخصيص نموذجية. وتُعد البرمجة الخطية (LP) أفضل تقنية معروفة في عائلة أدوات التحسين والتي تسمى البرمجة الرياضية؛ حيث تكون كافة العلاقات بين المتغيرات خطية في LP. كما يتم استخدامه على نطاق واسع في نظام الترصد الديموغرافي (انظر الحالة العملية ٦-٦). وتتميز نماذج البرمجة الخطية بأن لديها العديد من التطبيقات المهمة في الممارسة. وتتضمن هذه التطبيقات إدارة سلسلة التوريد، وقرارات مزج المنتجات، والتوجيه، وغير ذلك. ومن الممكن استخدام أشكال خاصة من النماذج لأجل تطبيقات محددة. فعلى سبيل المثال، تصف الحالة العملية (٦-٦) نموذج جدول البيانات الذي تم استخدامه لإنشاء جدول زمني للأطباء.

حالة عملية ٦-٦

مساعدة نموذج البرمجة العددية المختلطة لمركز جامعة تينيسي (Tennessee) الطبي في وضع جدول مواعيد الأطباء

إن مؤسسة الأطفال حديثي الولادة الإقليمية Regional Neonatal Associates هي عبارة عن مجموعة تضم تسعة أطباء يعملون في وحدة العناية المركزة لحديثي الولادة (NICU) في المركز الطبي لجامعة تينيسي (Tennessee) في نوكسفيل (Knoxville)، تينيسي (Tennessee). كما تخدم المجموعة أيضًا اثنتين من المستشفيات المحلية في

منطقة نوكســفيل (Knoxville) لأغراض الطوارئ. ولعدة سنوات، يقوم عضو واحد في المجموعة بوضع جدول مواعيد الأطباء يدويًا لعدة سنوات، بيد أنه، مع اقتراب تقاعده، كانت هناك حاجة إلى نظام أكثر تلقائية لوضع جدول مواعيد الأطباء. وقد رغب الأطباء في أن يكــون هــذا النظام قادرًا على موازنة عبء العمل الخاص بهم، حيث لم تســتطع الجداول الســابقة موازنة عبء العمل بينهم بشــكل مناسب. وبالإضافة إلى ذلك، فإن الجدول الزمني يحتاج لضمان تواجد الأطباء ٢٤ ساعة ٧ أيام في الأسبوع لتغطية وحدة العنايــة المركزة لحديثي الولادة NICU من قِبــل الأطباء، وإذا أمكن، يقوم هذا النظام باســتيعاب التفضيلات الفردية للأطبــاء لأنواع التناوب. ولمعالجة هذه المشــكلة، قام الأطباء بالاتصال بكلية العلوم الإدارية في جامعة تينيسي (Tennessee).

وقد تميزت مشــكلة جدولة مواعيد الأطباء إلى نوبات بأنها كانت تقوم على أســاس قيــود عــبء العمل وخيارات نمــط الحياة. وقد تمثلت الخطوة الأولى لحل مشــكلة في الجدولة وهي تجميع النوبات وفقًا لأنواعها (ليلًا ونهارًا). أما الخطوة التالية فقد تمثلت في تحديد قيود المشــكلة. ويحتاج النموذج لتغطية فترة تســعة أســابيع بتسعة أطباء، مــع اثنين من الأطباء يعملون طوال أيام الأســبوع وطبيب واحد لَيْلِي وفي عطلات نهاية الأســبوع، بالإضافة إلى ضرورة تعيين طبيب واحد فقط لتغطية ٢٤ ســاعة يوميًا لسبعة أيام في الأسبوع للمستشفيَين المحليَّين. كما تحتاج القيود الواضحة الأخرى أيضًا إلى النظر فيها. فعلى سبيل المثال، لا يمكن تكليف طبيب بوردية نهارية بعد قضائه لوردية ليلية.

المنهجية / الحل:

تم صياغة المشــكلة من خلال إنشــاء نموذج تحســين عددي مختلط. وقد قســم النموذج الأول حجم العمل بين تسعة أطباء بالتساوي. غير أن هذا النموذج لم يستطع أن يحدد عددًا متســاويًا من النوبات الليلية والنهارية بين الأطباء. وقد نشــأ عن ذلك مســألة التوزيع العادل. وبالإضافة إلى ذلــك، فقد كان للأطباء آراء مختلفة حول عبء العمل المحدد. وقد أراد ســتة أطباء وضع جدول زمني يتم فيه تعيين عددٍ متســاوٍ من النوبات النهارية والليلية لكل طبيب في الجدول الزمني الذي يستغرق تسعة أسابيع، في حــين يرغب الآخرون بوضع جدول زمني بناءً على تفضيل كل فرد من الأطباء للنوبات. ومن أجل ذلك تم تشــكيل نموذج جديد يمكنه تلبية متطلبات كلا الفريقين من الأطباء وسُــمِّي هذا النموذج بنموذج جدولة التفضيلات المختلطة (HPSM). ومن أجل تلبية متطلبات المســاواة للســتة أطباء الذين رغبوا في ذلك، قام النموذج أولًا بحساب حجم

العمل لمدة أســبوع واحد ثم قسمه لمدة تسـعة أسابيع بالنسبة لهم. وبهذه الطريقة، اســتطاع تقسيم العمل بالتساوي بين جميع الأطباء الستة، ثم قام بتقسيم عبء العمل للثلاثة أطباء المتبقين في الجدول الزمني لمدة تسـعة أسابيع وفقًا لتفضيلاتهم. ومن ثَم قام الأطباء بمراجعة الجدول الزمني الناتج فوجدوه أكثر قبولًا.

النتائج / الفوائد:

استطاعت طريقة جدولة التفضيلات المختلطة HPSM أن تستوعب كلًا من متطلبات المســاواة ومتطلبات الأفضلية الفردية للأطبــاء. وبالإضافة إلى ذلك، فقد تمكَّنت الجداول من هذا النموذج من تقديم أوقات راحة أفضل للأطباء مقارنةً بالجداول الزمنية السابقة، فضلًا عن إمكانية تلبية طلبات الإجازات في الجداول. كما يســتطيع نموذج HPSM حل مشاكل الجدولة المشابهة التي تتطلب التفضيلات النسبية بين أنواع النوبات.

وتســتطيع بعض التقنيات من مثل نماذج البرمجة العددية المختلطة أن تقوم بإنشــاء الجــداول المُثلى فضـلًا عن قدرتها على المســاعدة في العمليات. وقد تم اســتخدام هذه التقنيات في المؤسسات الكبيرة لفترة طويلة. والآن يمكن تنفيذ مثل هذه النماذج التحليلية التوضيحية في جداول البيانات وغيرها من البرامج المتاحة بسهولة.

أسئلة للمناقشة:

١- ما هي المشكلة التي واجهت مؤسسة الأطفال حديثي الولادة الإقليمية؟

٢- كيف استطاع نموذج HPSM تلبية جميع متطلبات الأطباء؟

Source: Adapted from Bowers, M. R., Noon, C. E., Wu, W., & Bass, J. K. (2016). Neonatal physician scheduling at the University of Tennessee Medical Center. Interfaces, 46(2), 168182-.

وعادةً ما تُظهر مشاكل توزيع البرمجة الخطية المميزات التالية:

- هناك كمية محدودة من الموارد الاقتصادية متاحة للتخصيص.
- تُستخدم الموارد في إنتاج المنتجات أو الخدمات.
- يوجد طريقتان أو أكثر يمكن من خلالها استخدام الموارد، ويسمى كل منها حل أو برنامج.
- ينتــج عن كل نشــاط (منتج أو خدمة) والذي يتم اســتخدام الموارد فيــه، عائدًا على الشروط الخاصة بالهدف المعلن.
- عادةً ما يكون التخصيص مقيدًا بعدة قيود ومتطلبات، تسمى القيود.

ويعتمد نموذج تخصيص البرمجة الخطية على الافتراضات الاقتصادية العقلانية التالية:

- يمكن مقارنة العوائد من التخصيصات المختلفة؛ بمعنى أنه يمكن قياسها بوحدة مشتركة (مثل: الدولار، والفائدة).
- العائد من أي تخصيص يكون مستقلًا عن التوزيعات الأخرى.
- إجمالي العائد هو مجموع العوائد الناتجة عن الأنشطة المختلفة.
- جميع البيانات معروفة على وجه التأكد.
- استخدام الموارد بطريقة أكثر اقتصادًا.

وعادةً ما تتضمن مشـاكل التخصيص عددًا كبيرًا من الحلـول الممكنة. واعتمادًا على الافتراضات الأساسـية، فإن عدد الحلول قـد يكون لا نهائي وقد يكون محدودًا، وعادةً مـا تنتج الحلول المختلفة مكافـآت مختلفة. ومن بين الحلول المتاحة يوجد على الأقـل حل واحد يعتبر هو الأفضل، بمعنى أن درجـة تحقيق الأهـداف المرتبطة به هي الأعلى (أي إن إجمالي المكافأة هـو الحد الأقصى)، وهذا ما يسمى الحل الأمثل، ويمكن العثور عليه باستخدام خوارزمية خاصة.

نموذج البرمجة الخطِّية:

يتكـون كل نموذج برمجة خطيـة LP من متغيرات القرار (والتي تكـون قيمتها غير معروفة ويتـم البحث عنها)، ودالة موضوعية (دالة حسـابية خطية تربط متغـيرات القرار بالهدف، كما تأخذ المعايير اللازمة لتحقق الأهداف، وكيفية تحسينها)، ومعاملات دالة الهدف (وحدة الربح أو معاملات التكلفة التي تشير إلى إسهام هدف وحدة واحدة من متغير القرار)، بالإضافة إلى القيود (المعبر عنها في شـكل عدم مسـاواة خطية أو مسـاواة تحد من الموارد و / أو المتطلبات؛ والتي تقوم بربط المتغيرات من خلال العلاقات الخطية)، والقدرات (والتي تصف الحدود العليا وأحيانًا الحدود الدنيا حسب القيود والمتغيرات)، ومعاملات المدخلات / المخرجات (التقنية) (والتي تشير إلى استخدام الموارد من أجل متغير القرار).

دعونا نتأمل هذا المثال: تحتاج شركة MBI، التي تقوم بصنع أجهزة حاسـب لأغراض خاصة، إلى اتخاذ قرار بخصوص: كمْ عدد أجهزة الحاسـب التي ينبغي إنتاجها الشـهر المُقبل في مصنع بوسـطن؟ حيث تدرس MBI نوعين من أجهزة الحاسـب: CC-7، والتي تتطلب ٣٠٠ يوم من العمـل و١٠,٠٠٠$ للخامات، وCC-8، والتي تتطلب ٥٠٠ يوم عمل و١٥,٠٠٠$ للخامات. وتبلغ قيمة المسـاهمة الربحية لكل جهاز CC-7 ٨,٠٠٠$، في حين تبلغ المسـاهمة الربحية لكل جهاز CC-8 ١٢,٠٠٠$. وتبلغ سـعة المصنع ٢٠٠,٠٠٠ يوم عمل في الشـهر، كما تبلغ الميزانية المادية ٨ مليون دولار شهريًّا. ويتطلب التسويق إنتاج ١٠٠ وحدة من CC-7 على الأقل بالإضافة إلى ما لا

يقل عن ٢٠٠ وحدة من CC-8 كل شــهر. وتكمن المشــكلة في زيادة أرباح الشركة إلى أقصى حد من خلال تحديد عدد وحدات CC-7 وعدد وحدات CC-8 التي ينبغي إنتاجها كل شهر. لاحظ أن هذا الأمر قد يســتغرق شــهورًا للحصول على بيانات مشــكلة معينة في البيئة الواقعية، وأثناء جمع البيانات سيكشف صانع القرار بلا شك الحقائق حول كيفية هيكلة النموذج المطلوب حله.

رؤية فنية ٦-١: البرمجة الخطِّية

ربما تكون البرمجة الخطية LP هي نموذج التحسين الأكثر شهرة؛ حيث يتعامل هذا النموذج مع التخصيص الأمثل للموارد بين الأنشــطة المتنافسة. وقد تم توضيح مشكلة التخصيص من خلال النموذج الموضح هنا.

وتكمن المشكلة في العثور على قيم متغيرات القرار س١ وس٢ وغيرها، بهدف وصول قيمة متغير النتيجة Z إلى أقصى حد ممكن، مع مراعاة مجموعة من القيود الخطية التي تعــبر عن كل من التقنية، وظروف الســوق، بالإضافة إلى المتغيرات الأخرى التي لا يمكن التحكــم فيهــا. أما العلاقات الرياضية فهي كل المعادلات الخطيــة بالإضافة إلى التباين. ومن الناحية النظرية، فإن أي مشــكلة تخصيص من هذا النوع لديها العديد من الحلول الممكنــة التي لا حصر لها. فمن خلال اســتخدام إجــراءات رياضية خاصة، يطبق منهج البرمجة الخطية LP إجراء بحث فريد يتم تنفيذها بواسطة الحاسب حيث يمكنه إيجاد أفضل حل (حلول) في غضون ثوانٍ، فضلًا عن قدرته على توفير تحليل الحساسية التلقائي.

النمذجة في البرمجة الخطِّية LP، مثال:

يمكــن تطوير نموذج برمجة خطية LP قياسي لمشــكلة شركة MPI التــي ذكرناها آنفًا. وكما ناقشنا في الفكرة التقنية (٦-١)، فإن نموذج البرمجة الخطية LP يحتوي على ثلاثة مكونات، وهي: متغيرات القرار، ومتغيرات النتائج، والمتغيرات التي لا يمكن التحكم فيها (القيود).

حيث تكون متغيرات القرار كما يلي:

$$X_1 = \text{units of CC} - 7 \text{ to be produced}$$
$$X_2 = \text{units of CC} - 8 \text{ to be produced}$$

ويكون متغير النتيجة كما يلي:

$$\text{Total profit} = Z$$

والهدف هو الوصول بالربح الإجمالي إلى أقصى حد ممكن:

$$Z = 8{,}000X_1 + 12{,}000X_2$$

وتكون المتغيرات (القيود) التي لا يمكن التحكم فيها كما يلي:

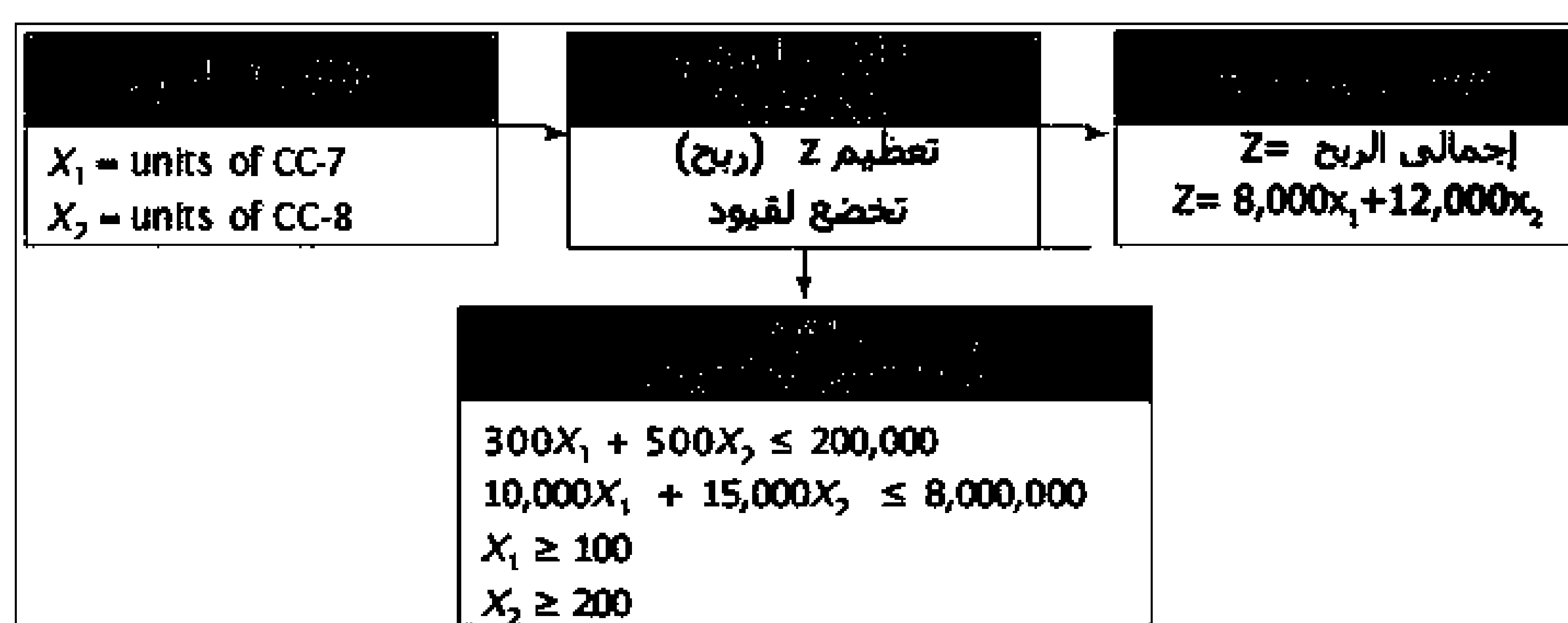
$$\text{Labor constraint: } 300X_1 + 500X_2 \leq 200{,}000 \text{ (in days)}$$

$$\text{Budget constraint: } 10{,}000X_1 + 15{,}000X_2 \leq 8{,}000{,}000 \text{ (in dollars)}$$

$$\text{Marketing requirement for CC} - 7\text{: } X_1 \geq 100 \text{ (in units)}$$

$$\text{Marketing requirement for CC} - 8\text{: } X_2 \geq 200 \text{ (in units)}$$

ويلخص الشكل (٦-٥) هذه المعلومات.

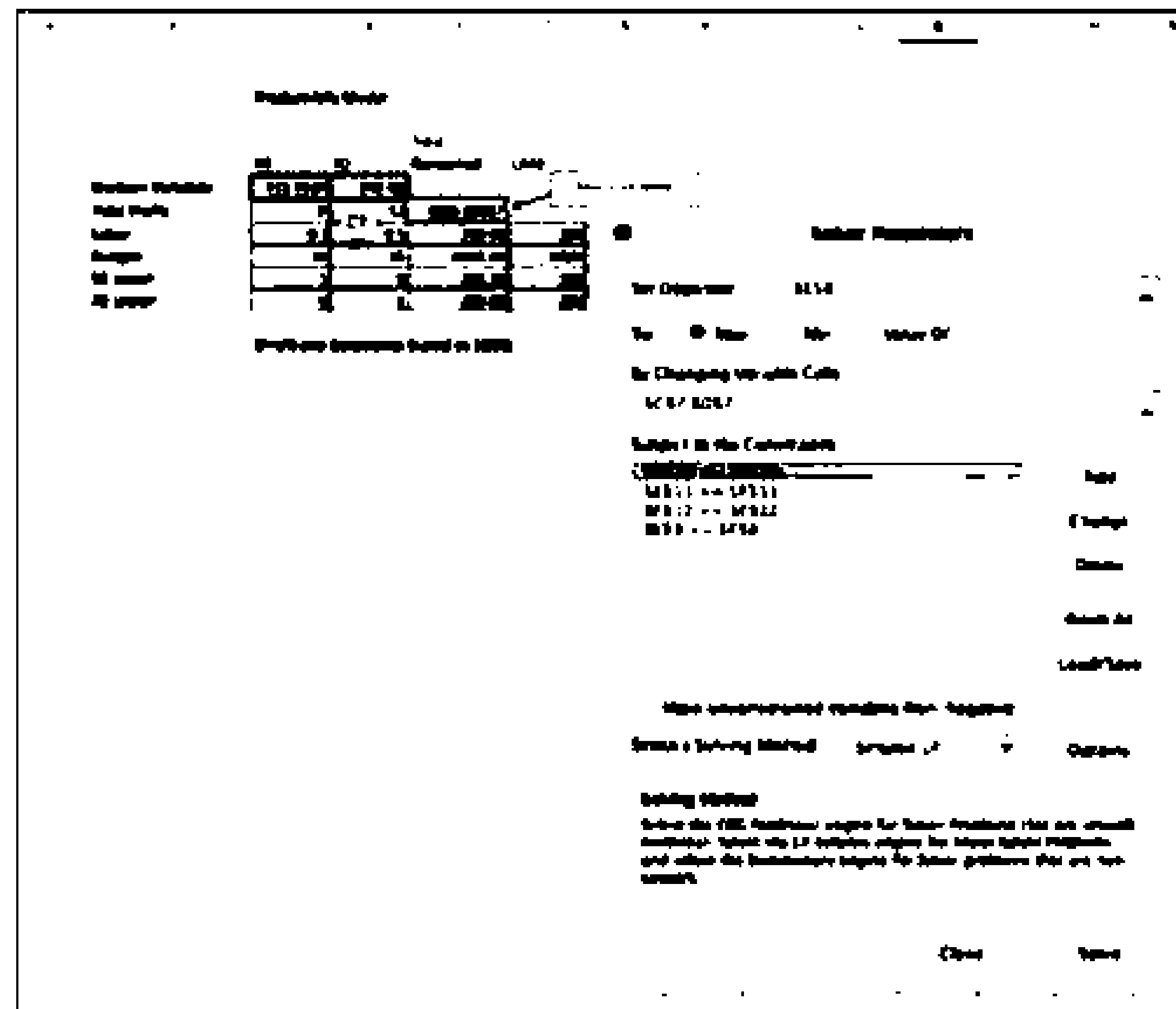


شكل ٦-٥: نموذج رياضي لمثال مزيج المنتج

وبالإضافة إلى المكونات التي ذُكِرت سابقًا، فإن النموذج يحتوي أيضًا على مكون رابع خفي؛ حيث يحتوي كل نموذج برمجة خطية LP على بعض المتغيرات الوسيطة الداخلية والتي لم تُذكر بشكل صريح. فعندما يكون الجانب الأيسر أقل دقة من الجانب الأيمن، فإن كلًا من قيود العمل والميزانية قد تسبب حدوث بعض الركود بهم. هذا الركود يتم تمثيله داخليًّا عن طريق متغيرات الركود التي تشير إلى وجود موارد زائدة متاحة. أما عندما يكون الجانب الأيسر ذا دقة أكبر من الجانب الأيمن فإن قيود متطلبات التسويق قد يكون لدى كل منها بعض الفائض، ويتم تمثيل هذا الفائض داخليًّا بواسطة المتغيرات الفائضة مما يدل على وجود بعض المساحة التي تتيح ضبط الجوانب اليمنى من هذه القيود. وتُعد كل من متغيرات الركود والفوائض متوسطة، كما أنها قد تكون ذات قيمة كبيرة لصانع القرار نظرًا لاستخدام طرق حل النماذج الخطية LP لها في وضع مَعْلَمات الحساسية لتحليلات "ماذا - لو" الاقتصادية.

ويحتــوي نموذج المزيج التســويقي على عدد لا نهائي من الحلول الممكنــة. فإذا افترضنا أن خطة الإنتــاج لا تقتــصر على الأعداد الصحيحة - وهو افتراض معقول في خطة الإنتاج الشــهرية - فإننا نريد حلًا يصل بإجمالي الربح إلى أقصى حد ممكن، ألا وهو الحل الأمثل. ولحسن الحظ يظهر برنامج Excel مع الأداة الإضافية Solver، والتي يمكنها الحصول بســهولة على الحل الأمثل (الأفضل) لهذه المشــكلة. وعــلى الرغــم من نقل موقع الوظيفة الإضافيــة Solver من أحد الإصــدارات إلى إصدار آخر، غير أنه مازال متاحًا كوظيفة إضافية مجانية، ويمكنك البحث عنه ضمن علامات تبويب البيانات وعلى شريط التحليل. فإنه يمكن تفعيله من خلال الانتقال إلى قائمة الخيارات في Excel وتحديد الوظائف الإضافية.

ونقــوم بإدخــال هذه البيانات مباشرة في جــدول بيانات Excel، ثم نعمل على تنشــيط Solver، ومن ثَم نقوم بتحديد الهدف (من خلال وضع الخلية المســتهدفة تســاوي الحد الأقصى)، بالإضافة إلى متغيرات القرار (عن طريق الإعداد بواســطة تغيير الخلايــا)، والقيود (من خلال التأكد من أن إجمالي العناصر المســتهلكة أقل من أو يســاوي الصفَّين الأوَّلين ويكون أكبر من أو يســاوي كلا الصفين الثالث والرابع). وتشكل كل من الخلايا C7 وD7 خلايا متغير القرار. وبعد تشغيل الوظيفة الإضافية Solver سيتم ملء النتائج في هذه الخلايا؛ حيث EV هي الخلية المستهدفة، والتي تُعد أيضًا المتغير الناتج، كما أنها تمثل منتجًا من خلايا متغير القرار ومعاملات الربح لكل وحدة (في الخلايا C8 وD8). لاحظ أنه تم قســمة جميع الأرقام على ١٠٠٠ لجعل كتابتها أسهل (باستثناء متغيرات القرار). وتصف الصفوف من ٩ إلى ١٢ قيود المشــكلة: القيود على قدرة اليد العاملة، وعلى الميزانية، بالإضافة إلى القيود على الحد الأدنى المطلوب لإنتاج المنتجين X_1 و X_2. ويحدد كلا العمودين C وD معاملات هذه القيود. وينطوي العمــود E على الصيغ التي تضاعف متغيرات القرار (خلايــا C7 وD7) مع معاملات كل منهما في كل صــف. أما القيمة الجانبيــة لهذه القيود فيحددها العمود F. ولتطوير مضاعفات الصفوف والأعمدة بسهولة، فإنه يمكن استخدام إمكانيات مصفوفة إكسل (على سبيل المثال، دالة SUMPRODUCT). وبعد إعداد حســابات النموذج في Excel، فإن هذا هو الوقت المناســب لاســتدعاء الوظيفة الإضافية Solver. ويؤدي النقر فوق الوظيفة الإضافية لـ Solver (مرة أخرى ضمن مجموعة Analysis ضمن علامــة تبويــب البيانات) إلى فتح مربع حوار (إطار) يتيح لك تحديــد الخلايا أو النطاقات التي تحدد خلية دالة الهدف ومتغيرات القرار / التغيير (الخلايا) والقيود. كما نقوم أيضًا باختيار طريقة الحل من الاختيارات (عادةً البرمجة الخطية البسيطة LP)، ومن ثَم نقوم بحل المشكلة. وبعد ذلك نقوم باختيار التقارير الثلاثة - الإجابة والحساســية والقيود - للحصول على الحل الأمثل من $5.066.667 = الربح، $X_2 = 200$، $X_1 = 333.33$، كما يتضح من الشكل (٦-٦). حيث تنتج Solver ثلاثة تقارير مفيدة حول الحل. يمكنك القيام بتجربتها. وقد أصبح الآنSolver ذا قدرة على حل مشــاكل البرمجة غير الخطية ومشاكل البرمجة الصحيحة باستخدام أساليب الحل الأخرى المتوفرة به.



شكل ٦-٦: مثال لحل Solver من Excel لمزيج المنتجات

وفيما يلي مثال لتوضيح مدى قوة نمذجة جداول بيانات دعم القرار، وقد تم وضع هذا المثال من قِبل البروفيسور Rick Wilson من جامعة أوكلاهوما الحكومية.

	A	B	C	D	E	F	G	H
1								
2				Electoral			Influence	
3			State	Votes	W/E	N/S	Function	
4			NV	6	West		F1	
5			CO	9	West		F2	
6			IA	6	West	North	F3	
7			WI	10	West	North	F1	
8			OH	18	East	North	F2	
9			VA	13	East	South	F3	
10			NC	15	East	South	F1	
11			FL	29	East	South	F3	
12			NH	4	East		F1	
13								
14		F1	Young	Old				
15		Men	3	1	4			
16		Women	3	3	6			
17			6	4	10	Total		
18								
19		F2	Young	Old				
20		Men	1.5	2.5	4			
21		Women	2.5	1	3.5			
22			4	3.5	7.5	Total		
23								
24		F3	Young	Old				
25		Men	2.5	2.5	5			
26		Women	1	2	3			
27			3.5	4.5	8	Total		
28								

شكل ٦-٧: مثال لبيانات توزيع موارد الانتخابات

حيث يصف الجدول الوارد في الشكل (٦-٧) بعض البيانات الافتراضية وسمات تسع «ولايات متقلبة» لانتخابات عام ٢٠١٦. حيث تشمل سمات التسع ولايات مَن كان عددهم من الأصوات الانتخابية، واثنتين من الواصفات الإقليمية (لاحظ أن هناك ثلاث ولايات تصنَّف على أنها ليست شمالية ولا جنوبية)، فضلًا عن "وظيفة التأثير" المقدرة والتي تتعلق بزيادة دعم المرشح لكل وحدة من حملات الاستثمار المالي في تلك الحالة. فعلى سبيل المثال، تُبين دالة التأثير F1 أنه سيكون هناك زيادة إجمالية قدرها ١٠ وحدات في دعم الناخبين (تظل الوحدات العامة هنا كما هي عامة) بالنسبة لكل وحدة مالية مستثمرة في تلك الولاية، كما تتكون من زيادة في دعم الشباب بمقدار ثلاث وحدات، ودعم كبار السن من الرجال بمقدار وحدة واحدة، والنساء صغار السن والمسنات بمقدار ٣ وحدات لكل منهم.

وتضم الحملة ١٠٥٠ وحدة مالية للاستثمار في الولايات التسع؛ حيث ينبغي ألا يقل نصيب كل ولاية من الاستثمار عن ٥٪ من إجمالي الاستثمارات الكلية، كما ينبغي ألا يزيد عن ٢٥٪ من إجمالي الاستثمارات المستثمرة في أي من الولايات. ولا يلزم استثمار جميع الوحدات البالغ عددها ١٠٥٠ وحدة (ينبغي أن يتعامل النموذج الخاص بك مع هذا بشكل صحيح).

أيضًا، فإن الحملة لديها بعض القيود الأخرى. ومن وجهة نظر الاستثمار المالي، فإنه ينبغي أن يكون لدى الولايات الغربية (إجمالًا) استثمارات في الحملات بمستوى لا يقل عن ٦٠٪ من إجمالي المستثَمر في الولايات الشرقية. أما من حيث الأشخاص المتأثرين، فإنه ينبغي أن يؤدي قرار تخصيص الاستثمارات المالية للولايات إلى مالا يقل عن ٩٢٠٠ من إجمالي المتأثرين. وبشكل عام، فإنه ينبغي أن يكون العدد الإجمالي للإناث المتأثرين أكبر من أو يساوي العدد الإجمالي للذكور المتأثرين، بالإضافة إلى أنه يجب أن يكون ٤٦٪ على الأقل من مجموع الأشخاص المتأثرين من «المسنين».

وتتمثل مهمتنا في إنشاء نموذج برمجة صحيح مناسب يحدد العدد الصحيح الأمثل (أي العدد الكامل) للوحدات المالية الخاصة بالولايات والتي ترفع من مجموع منتجات وحدات التصويت الانتخابية، التي يتم استثمارها تبعًا للقيود الأخرى المذكورة آنفًا إلى أقصى حد ممكن. (وهكذا، وبشكل غير مباشر، فإن هذا النموذج يفضل الولايات التي تتمتع بعدد كبير من الأصوات الانتخابية). لاحظ أنه يجب أن تؤدي جميع القرارات الخاصة بالتخصيص في النموذج إلى قيم صحيحة، حتى يستطيع موظفو الحملة التنفيذ بسهولة.

واعتمادًا على الأسئلة التالية التي يجيبون عنها، يمكن تصنيف الجوانب الثلاثة الخاصة بالنماذج، وتتمثل هذه الأسئلة في:

١- مـا الـذي نتحكم فيه؟ المبلغ المسـتثمر في الإعلانات عبر التسـع ولايات، نيفـادا Nevada، وكولـورادو Colorado، وأيوا Iowa، وويسكونسـن Wisconsin، وأوهايو Ohio، وفرجينيا Virginia، بالإضافة إلى نورث كارولينا North Carolina، وفلوريدا Florida، ونيو هامبشاير New Hampshire، والتـي تُمثلهـا متغيرات القرار التسـعة: NV، وCO، وIA، وWI OH، وVA، وNC، وFL، بالإضافة إلى NH.

٢- مـاذا نريـد أن نحقق؟ نريد تحقيق أقـصى زيادة ممكنة للعدد الإجمالي لمكسـب الأصوات الانتخابيـة. فنحن على درايـة بقيمة كل صوت انتخابي في كل ولاية (EV)، حيث يعادل هذا الصوت الانتخابي *EV الاستثمارات المجمعة للولايات التسع، أي:

$$Max\ (6NV + 9CO + 6IA + 10WI + 18OH + 13VA + 15NC + 29FL + 4NH)$$

٣- ما الذي يقيدنا؟

وفيما يلي القيود كما ذُكِرت في وصف المشكلة:

(أ) مالا يزيد عن ١,٠٥٠ وحدة مالية للاستثمار فيها، أي:

$NV + CO + IA + WI + OH + VA + NC + FL + NH < = 1,050$.

(ب) استثمار ٥٪ على الأقل من الإجمالي في كل ولاية، أي:

$$NV > = 0.05\ (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$
$$CO > = 0.05\ (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$
$$IA > = 0.05\ (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$
$$WI > = 0.05\ (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$
$$OH > = 0.05\ (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$
$$VA > = 0.05\ (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$
$$NC > = 0.05\ (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$
$$FL > = 0.05\ (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$
$$NH > = 0.05\ (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

ويمكننا تنفيذ هذه القيود التسع بعدة طرق عن طريق استخدام Excel.

(ج) لا يزيد الاستثمار عن ٢٥٪ من الإجمالي في كل ولاية.

وكما هو الحال في (ب)، فإننا نحتاج مرة أخرى إلى تسع قيود فردية لأننا لا نعرف كم من ١,٠٥٠ سوف نستثمر. كما يجب أن تكون القيود بعبارات «عامة».

NV < = 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)
CO < = 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)
IA < = 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)
WI < = 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)
OH < = 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)
VA < = 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)
NC < = 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)
FL < = 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)
NH < = 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)

(د) يجب ألا تقل مستويات استثمار الولايات الغربية عن ٦٠٪ من مستويات الولايات الشرقية.

West States = NV + CO + IA + WI
East States = OH + VA + NC + FL + NH

So, (NV + CO + IA + WI) > = 0.60 (OH + VA + NC + FL + NH). Again we can

ويتم تنفيذ هذا القيد من خلال طرق متنوعة باستخدام Excel.

(هـ) تأثير مالا يقل عن ٩٢٠٠ من مجموع الناس، ويكون كالتالي:

(10NV + 7.5CO + 8IA + 10WI + 7.5OH + 7.5VA + 10 NC + 8FL + 8 NH) > = 9,200

(و) أن يكون تأثير عدد الإناث على الأقل مثل الذكور، مما يتطلب انتقال وظائف التأثير.

F1 = 6 women influenced, F2 = 3.5 women
F3 = 3 women influenced
F1 = 4 men influenced, F2 = 4 men
F3 = 5 men influenced

فإذا كان إنجاز الإناث > = الذكور، فإننا نحصل على:

(6NV + 3.5CO + 3IA + 6WI + 3.5OH + 3.5VA + 6NC + 3FL + 3NH) > =
(4NV + 4CO + 5IA + 4WI + 4OH + 4VA + 4NC + 5FL + 5NH)

وكما هو الحال من قبل، فإنه يمكننا تنفيذ ذلك في Excel بطريقتين مختلفتين.

(ز) ينبغي ألا تقل نسبة المسنين عن ٤٦٪ من المجموع الكلي للأشخاص.

وقد كان جميع الأشـخاص المتأثرين على الجانب الأيسر من القيد (e). ولذلك فإن المسنين المتأثرين سوف يكونون كالتالي:

(4NV + 3.5CO + 4.5IA + 4WI + 3.5OH + 3.5VA + 4NC + 4.5FL + 4.5NH)
This would be set >= 0.46* the left-hand side of constraint (e). (10NV + 7.5CO + 8IA + 10WI + 7.5OH + 7.5VA +10NC + 8FL + 8NH), which would give a right-hand side of
(0.46NV + 3.45CO + 3.68IA + 4.6WI + 3.45OH + 3.45VA + 4.6NC + 3.68FL + 3.68NH)

هذا هو القيد الأخير إلى جانب فرض أن تكون جميع المتغيرات أعدادًا صحيحة.

وقـد تـم ذكر كل شيء في الشروط الجبرية، وسـيكون لهذا النمـوذج الصحيح من البرمجة ٩ متغيرات قرار و٢٤ قيدًا (قيد واحد للمتطلبات الصحيحة).

التنفيذ:

سـيكون أحد المناهج صارمًا في تنفيذ النموذج «النموذج القياسي»، أو نموذج شـكل الصف، حيـث تتم كتابة جميع القيـود بمتغيرات القرار على الجانب الأيسر، وتتم كتابة رقم على الجانب الأيمن. ويوضح الشكل (٦-٨) مثل هذا التنفيذ، بالإضافة إلى أنه يعرض النموذج الذي تم حله.

شكل ٦-٨: نموذج لتوزيع موارد الانتخابات - إصدار قياسي

ويمكننا بدلًا من ذلك اســتخدام جدول البيانات لحساب أجزاء مختلفة من النموذج بطريقة أقل صرامة، وكذلك تنفيذ القيود المتكررة (b) و(c) بشكل فريد، بالإضافة إلى الحصول على جدول بيانات أكثر اختصارًا (غير أنها ليست واضحة). وهذا واضح في الشكل (٦-٩).

شكل ٦-٩: معادلة مضغوطة لتوزيع موارد الانتخابات

كـما يمكـن أيضًا تحديد نماذج البرمجة الخطيـة LP (وتخصصاتها وتعميماتهـا) مباشرةً في عدد من أنظمة النمذجة الأخرى سهلة الاستخدام. ومن أشهر هذه النماذج: نموذج ليندو Lindo، ولينجو Lindo Lingo (Systems, Inc., lindo.com; demos are available)، حيث Lindo هو نموذج برمجة خطية LP كما أنه نظام برمجة صحيح. ويتم تحديد النماذج في الأساس بالطريقة ذاتها التي يتم تحديدها بها جبريًّا. وقد اسـتندت الشركة على نجاح ليندو Lindo في تطوير لغة لينجو Lingo، وهي لغة النمذجة التي تنطوي على محسن ليندو Lindo القوي والملحقات لحل المشاكل غير الخطية. وهناك العديد من لغات النمذجة الأخرى المتوفرة مثل: AMPL، وAIMMS، بالإضافة إلى MPL، وXPRESS، وغيرها.

ومن خلال مجموعة متنوعة من أسـاليب البرمجة الرياضية، يمكن حل نماذج التحسـين الأكثر شيوعًا، وتتضمن هذه الأساليب ما يلي:

- التكليف (أفضل مطابقة للأهداف).
- البرمجة الحركية.
- برمجة الأهداف.
- الاستثمار (زيادة معدل العائد).
- البرمجة الخطية والصحيحة.
- نماذج الشبكة للتخطيط والجدولة.
- البرمجة غير الخطية.
- الاستبدال (وضع ميزانية رأس المال).
- النقل (تقليل تكلفة الشحنات).
- نماذج المخزون البسيط (مثل: كمية الطلب الاقتصادية).

أسئلة مراجعة على القسم ٦-٦:

١- اذكر مع التوضيح الافتراضات المتضمنة في البرمجة الخطية LP.

٢- اذكر مع التوضيح خصائص البرمجة الخطية.

٣- قم بوصف مشكلة التخصيص.

٦-٧ الأهداف المتعددة، وتحليل الحساسية، وتحليل «ماذا - لو»، والبحث عن الهدف:

تنطوي الكثير من حالات اتخاذ القرار إن لم يكن معظمها، على التلاعب بين الأهداف المتنافسة والبدائـل. وبالإضافـة إلى ذلك، فإن هناك عـدم تأكد بدرجة كبيرة حـول الافتراضات والتنبؤات المسـتخدمة في بناء نموذج التحليلات التوجيهية. وتقر الفقرات التالية ببساطة أنه يتم معالجتها أيضًا في برامج وتقنيات التحليلات التوجيهية. كما أن تغطية هذه التقنيات عادةً ما تكون شائعةً في الدورات التحليلية التوضيحية أو بحوث العمليات / العلوم الإدارية.

الأهداف المتعددة:

يتم تحليل القرارات الإدارية بهدف التقييم قدر الإمكان، لمدى مساعدة كل بديل للمديرين على التقدم نحو أهدافهم. وللأسف، فإن تقييم المشاكل الإدارية نادرًا ما يتم بهدف واحد بسيط، مثل: تعظيم الربح. وتُعد نظم إدارة اليوم معقدة بدرجة كبيرة، ونادرًا ما تجد أحد هذه الأنظمة لديها هدف واحد فقط، حيث يرغب المديرون بدلًا من ذلك في تحقيق أهداف متزامنة في وقت واحد، والتي قد تتعارض مع بعضها البعض. الجدير بالذكر أن أصحاب المصلحة المختلفة تكون لديهم أهداف مختلفة. لذلك، فإنه غالبًا ما يكون من الضروري تحليل كل بديل في ضوء تحديد كل هدف من الأهداف المتعددة (انظر Koksalan & Zionts, 2001).

وعلى سبيل المثال، فلنعتبر أن هناك شركة تحقق أرباحًا، فإنه بالإضافة إلى هدف الشركة في تحقيق مكاسب مالية، فإن لديها أهدافًا أخرى ترغب في تحقيقها وهي النمو وتطوير منتجاتها وموظفيها، وتوفير الأمن الوظيفي لعمالها، وخدمة المجتمع. كما يرغب المديرون في إرضاء المساهمين وفي الوقت ذاته يرغبون في التمتع برواتب عالية وحسابات مصرفية.وبالإضافة إلى ذلك، فإن الموظفين يرغبون في زيادة أجورهم واستحقاقاتهم من الأرباح. وعند اتخاذ قرار - بخصوص مشروع استثماري، على سبيل المثال - فإن بعض هذه الأهداف تكمل بعضها البعض، في حين يتعارض بعضها الآخر. وقد وصف Kearns (2004) كيف تقوم عملية التحليل الهرمي (AHP) جنبًا إلى جنب مع البرمجة الصحيحة، بمعالجة عدة أهداف في تقييم استثمارات تقنية المعلومات (IT).

وتستند العديد من النماذج الكمية لنظرية القرار على مقارنة مقياس واحد للفعالية.وبشكل عام، فإن ذلك يُعد شكلًا من أشكال المنفعة بالنسبة لصانع القرار. ولذلك، فإنه من الضروري عادةً تحويل مشكلة ذات أهداف متعددة إلى مشكلة قياس فعالية واحدة قبل مقارنة آثار الحلول. وتُعد هذه طريقة شائعة لمعالجة أهداف متعددة في نموذج LP.

وهناك بعض الصعوبات التي قد تنشأ عند تحليل الأهداف المتعددة، وتتمثل في:

- صعوبة الحصول عادةً على بيان صريح لأهداف المنظمة.
- قد يقوم صانع القرار بتغيير الأهمية الخاصة بأهداف معينة مع مرور الوقت أو من أجل سيناريوهات القرار المختلفة.
- النظر إلى الأهداف والمناهج الفرعية بنظرة مختلفة على مستويات مختلفة من التنظيم وداخل الإدارات المختلفة.
- الاستجابة للتغيرات في المنظمة وبيئتها مما يترتب عليه تغيير الأهداف.

- قد يصعب تحديد العلاقة بين البدائل ودورها في تحديد الأهداف.
- حل المشـاكل المعقدة من قِبل مجموعات من صانعي القرار، لكل واحد منهم رؤية شـخصية خاصة به في تناول المشكلة.
- قيام المشاركين بتقييم الأهمية (الأولويات) الخاصة بمختلف الأهداف، بطريقة مختلفة.

وعند العمل مع مثل هذه الحالات، فإن هناك عدة طرق يمكن اسـتخدامها لمعالجة الأهداف المتعددة، وتتمثل أكثر هذه الطرق شيوعًا في:

- نظرية المنفعة.
- برمجة الأهداف.
- التعبير عن الأهداف كقيود، باستخدام LP.
- نظام النقاط.

تحليل الحساسية:

يقوم منشـئ النماذج بعمل تنبؤات وافتراضات تتعلق ببيانات المدخلات، والتي يتناول الكثير منهـا تقييم العقود الآجلة غير المؤكـدة. وعندما يتم حل النموذج، فإن النتائج تعتمد على هذه البيانات. ويحاول تحليل الحساسية تقييم أثر التغيير في بيانات المدخلات أو المَعْلمات على الحل المقترح (أي: متغير النتيجة).

ونظرًا لأن تحليل الحساسية يسمح بالمرونة والتكيف مع الظروف المتغيرة ومتطلبات مختلف حـالات اتخاذ القرار، كما أنه يوفر فهمًا أفضل للنموذج ووضـع صنع القرار الذي يحاول وصفه، ويسـمح للمدير بإدخـال البيانات لزيادة الثقة في النموذج، فإنه يُعـد مُهمًّا للغاية في التحليلات التوجيهية. ويقوم تحليل الحساسية باختبار العلاقات مثل:

- أثر التغيرات في المتغيرات الخارجية (الغير قابلة للسـيطرة عليها) والمَعْلمات parameters على متغير النتيجة (النتائج).
- أثر عدم التأكد في تقدير المتغيرات الخارجية.
- تأثيرات التفاعلات المستقلة المختلفة بين المتغيرات.
- قوة القرارات في ظل الظروف المتغيرة.

وتستخدم تحليلات الحساسية من أجل:

- مراجعة النماذج لاستبعاد الحساسيات الكبيرة جدًا.

- إضافة تفاصيل حول المتغيرات أو السيناريوهات الحساسة.
- الحصول على تقديرات أفضل للمتغيرات الخارجية الحساسة.
- تغيير نظام العالم الحقيقي للحد من الحساسيات الفعلية.
- قبول واستخدام العالم الحقيقي الحساس (وبالتالي سريع التأثر بالنقد)، مما يؤدي إلى المراقبة المستمرة والوثيقة للنتائج الفعلية.

أما نوعا تحليلات الحساسية فهما يتمثلان في التلقائية، والتجربة والخطأ.

- تحليل الحساسية التلقائي: يتم إجراء الحساسية التلقائي في تطبيقات نموذجية كمية قياسية مثل LP. وعلي سبيل المثال، فإن هذا التحليل يقوم بالإبلاغ عن النطاق الذي يمكن أن يختلف فيه متغير إدخال معين أو قيمة مَعْلمة (مثل: تكلفة الوحدة) دون أن يكون له أي تأثير خطير على الحل المقترح. وعادةً ما يقتصر تحليل الحساسية التلقائي على تغيير واحد في كل مرة، وليس لكل المتغيرات بل فقط لمتغيرات معينة. وعلى الرغم من ذلك، فإن هذا التحليل قوي بسبب قدرته على إنشاء نطاقات وحدود بسرعة كبيرة (وبجهد حاسوبي إضافي أو بدون أي جهد إضافي). ويقوم Silver وتقريبًا جميع حزم البرامج الأخرى مثل Lindo بتوفير تحليل الحساسية. وبالنظر في مثال شركة MBI الذي قدمناه فيما سبق. فإنه من الممكن استخدام تحليل الحساسية لتحديد ما إذا كان الجانب الأيمن من القيد التسويقي على CC-8 بواسطة وحدة واحدة، فإن صافي الربح سيزيد بمقدار ١,٣٣٣.٣٣ $. وهذا التناقص إلى الصفر يكون صحيحًا بالنسبة إلى الجانب الأيمن. كما يمكن إجراء تحليل إضافي كبير على طول هذه الخطوط.
- تحليل الحساسية والكشف عن الأخطاء: من خلال نهج التجربة والخطأ البسيط يمكن تحديد تأثير التغييرات في أي متغير، أو في عدة متغيرات؛ حيث يمكنك تغيير بعض بيانات الإدخال وحل المشكلة مرة أخرى. وبتكرار التغييرات عدة مرات، فإنه قد يتم اكتشاف حلول أفضل وأحسن. ومثل هذا الاختبار، والذي يسهل تنفيذه عند استخدام برامج النمذجة المناسبة، مثل: برنامج Excel، له أسلوبان: تحليل "ماذا - لو"، بالإضافة إلى البحث عن الهدف Goal Seeking.

تحليل "ماذا - لو":

وقد أُنشئ تحليل "ماذا - لو" على أساس سؤال ماذا سيحدث للحل إذا تم تغيير متغير إدخال أو افتراض أو قيمة مَعْلمة؟ وفيما يلي بعض الأمثلة:

- ماذا سيحدث لإجمالي تكلفة المخزون إذا ارتفعت تكلفة حفظ المخزون بنسبة ١٠٪؟
- ما هي حصة السوق إذا ارتفعت ميزانية الإعلان بنسبة ٥٪؟

ويُعد طرح هذه الأنواع من الأسـئلة على نموذج الحاسـب والحصول على إجابات فورية أمرًا سـهلًا بالنسـبة للمديرين، إذا ما تعاملوا مع واجهة المستخدم المناسـبة. وعلاوةً على ذلك، فإن المديرين يمكنهم إجراء العديد من الحالات، وبالتالي تغيير النسـبة المئوية، أو أي بيانات أخرى في السؤال، حسب رغبتهم. ويستطيع صانع القرار فعل كل ذلك بشكل مباشر، دون مبرمج حاسب.

ويقدم الشـكل (٦-١٠) مثالًا لجدول بيانات عن سـؤال "ماذا - لو" لمشكلة التدفق النقدي. فعندمـا يقوم المسـتخدم بتغيير الخلايا التـي تحتوي على المبيعات الأولية (مـن ١٠٠ إلى ١٢٠) ومعدل نمو المبيعات (من٣٪ إلى ٤٪ لكل ربع)، فإن البرنامج يقوم على الفور بإعادة حساب قيمة صافي الربح السنوي (من ١٢٧$ إلى ١٨٢$). وقد كانت المبيعات الأولية في البداية ١٠٠، حيث نمت بمعدل ٣٪ لكل ربع، محققةً ربحًا صافيًا قدره ١٢٧$. حيث يؤدي تغيير خلية المبيعات الأولية إلى ١٢٠ بالإضافة إلى تغيير معدل نمو المبيعات إلى ٤٪ إلى رفع صافي الربح السـنوي إلى ١٨٢$. ويُعد تحليل "ماذا - لو" شـائعًا في العديـد من أنظمة القرار. ويتم منح المسـتخدمين الفرصة لتغيير إجاباتهم على بعض الأسئلة الخاصة بالنظام، والحصول على اقتراح مُعَدَّل.

4						
5						
6						
7	Unit revenue	$ 1.20				
8	Unit cost	$ 0.60				
9						
10	Initial sales	120				
11	Sales growth rate	0.04				
12						
13	Annual net profit	$ 182				
14						
15						
16						
17			Cash Flow Model for 1996			
18						Annual
19		Qtr1	Qtr2	Qtr3	Qtr4	Total
20	Sales	120	125	130	135	510
21	Revenue	$ 144	$ 150	$ 156	$ 162	$ 611
22	Variable cost	$ 72	$ 75	$ 78	$ 81	$ 306
23	Fixed cost	$ 30	$ 31	$ 31	$ 32	$ 124
24	Net profit	$ 42	$ 44	$ 47	$ 49	$ 182
25						

تقييم المبيعات الأولية (الخلية B10)
ومعدل نمو المبيعات (الخلية B11)
لتقييم التغير في الأرباح السنوية

مبيعات أولية بمقدار 100 بمعدل نمو ربع سنوي
بمقدار 3% ينتج صافي ربح سنوي قدره 127
دولار
بالمقارنة مع حالة ماذا-لو لمبيعات أولية بمقدار 120
بمعدل نمو ربع سنوي قدره 4%

شكل ٦-١٠: نموذج لحالة "ماذا - لو" من خلال ورقة عمل

البحث عن الهدف Goal Seeking:

يقوم البحث عن الهدف Goal Seeking بحساب قيم المدخلات اللازمة لتحقيق المستوى المطلوب من المخرجات (الهدف). ويمثل هذا النهج نهج الحل المتأخر. وفيما يلي بعض الأمثلة على البحث عن الهدف Goal Seeking:

- ما هي الميزانية السنوية للبحث والتطوير اللازمة لتحقيق معدل نمو سنوي بنسبة ١٥٪ بحلول عام ٢٠١٨؟
- كم عدد الممرضات اللازمة لتقليل متوسط وقت الانتظار للمريض في غرفة الطوارئ إلى أقل من ١٠ دقائق؟

ويقدم الشكل (٦-١١) مثالًا على البحث عن الهدف Goal Seeking. فعلى سبيل المثال، معدل العائد الداخلي (IRR) في نموذج التخطيط المالي Excel هو معدل الفائدة الذي ينتج قيمة صافية حالية (NPV) صفر. ويمكننا حساب صافي القيمة الحالية للاستثمارات المخطط لها، من خلال النظر إلى تدفق العائدات السنوية في العمود E. كما يمكننا تحديد معدل العائد الداخلي من خلال تطبيق البحث عن الهدف Goal Seeking، حيث تكون NPV صفرًا. أما الهدف المراد تحقيقه والذي يحدد معدل العائد الداخلي لهذا التدفق النقدي، بما في ذلك الاستثمار؛ فهو أن تساوي القيمة الصافية الحالية NPV صفرًا. وقد قمنا بتعيين خلية NPV إلى القيمة ٠ عن طريق تغيير خلية معدل الفائدة. وكان الجواب ٣٨,٧٧٠٥٩٪.

5				
6				
7	Investment Problem		Initial Investment:	$ 1,000.00
8	Example of GoalSeeking		Interest Rate:	10%
9				
10	Find the Interest Rate		Annual	NPV
11	(the Internal Rate of	Year	Returns	Calculations
12	Return-IRR)	1	$ 120.00	$109.09
13	that yields an NPV	2	$ 130.00	$118.18
14	of $0	3	$ 140.00	$127.27
15		4	$ 150.00	$136.36
16		5	$ 160.00	$145.45
17		6	$ 152.00	$138.18
18		7	$ 144.40	$131.27
19		8	$ 137.18	$124.71
20		9	$ 130.32	$118.47
21		10	$ 123.80	$112.55
22				
23			The NPV Solutions:	$261.55
24				

شكل ٦-١١: تحليل البحث عن الهدف

حساب نقطة تحقيق التعادل عن طريق استخدام البحث عن الهدف Goal Seeking:

تســتطيع بعض حزم برامج تصميم النماذج أن تقوم بحســاب نقاط التعادل بشــكل مباشر، والتي تُعد تطبيقًا هامًا للاســتهداف Goal Seeking. وهذا يتضمن قيمة متغيرات القرار (مثل: الكمية التي يُراد إنتاجها) التي تحقق ربحًا صفريًا.

وقد يكون من الصعب إجراء تحليل الحساسية في العديد من برامج التطبيقات العامة، نظرًا لأن الروتين المكتوب مســبقًا لا يمثل عادةً ســوى فرصة محدودة لطرح أسئلة "ماذا - لو". وينبغي أن يكون تنفيذ كل من "ماذا - لو"، والبحث عن الهدف Goal Seeking أمرًا سهلًا في نظام الترصد الديموغرافي (DSS).

أسئلة مراجعة على القسم ٦-٧:

١- اذكر بعضًا من الصعوبات التي قد تظهر عند تحليل أهداف متعددة.

٢- اذكر أسباب القيام بتحليل الحساسية.

٣- وضح لماذا قد يستخدم أحد المديرين تحليل "ماذا - لو".

٤- وضح لماذا قد يقوم المدير باستخدام البحث عن الهدف Goal Seeking.

٦-٨ تحليل القرار باستخدام جداول القرار وأشجار القرار:

بالنســبة لحــالات القــرار التــي تنطــوي عــلى عدد محــدود مــن البدائــل والــذي عادةً ما يكون عــددًا غــير كبير، فإنــه يتم تخطيطهــا باســتخدام نهج يســمى تحليل القــرار (انظر .(Arsham, 2006a,b; Decision Analysis Society, decision analysis. society. informs. org حيث يتم باســتخدام هذا النهج، سرد البدائل في جدول أو رســم بياني، مع مســاهماتها المتوقعة للهدف (الأهداف) بالإضافة إلى احتمالية الحصول على المســاهمة. ومن الممكن تقييم ذلك من أجل تحديد البديل الأفضل.

ويمكن اســتخدام جداول القرارات أو أشــجار القرار لصياغة حــالات الأهداف الفردية. كما يمكن وضع أهداف متعددة (معايير) مع العديد من التقنيات الأخرى، الموضحة لاحقًا في هذا الفصل.

جداول القرارات:

تقــوم جداول القرارات بتنظيم المعلومات والمعرفة بشــكل ملائم بطريقة منتظمة وجدولية بغرض إعدادها للتحليل. فعلى ســبيل المثال، لنفترض أن شركة استثمار تقوم بدراسة الاستثمار في

بديل من البدائل الثلاثة: السـندات، أو الأسـهم، أو شهادات الإيداع (CDs)، فإن اهتمام الشركة سـيكون مُنْصبًّا عـلى هدف واحد، ألا وهو زيـادة العائد على الاسـتثمار إلى أقصى حد بعد عام واحد. فإذا اهتمت الشركة بأهداف أخرى، مثل السلامة أو السيولة، فإنه سيتم تصنيف المشكلة باعتبارها أحد مشاكل تحليل القرارات متعدد المعايير (انظر Koksalan & Zionts, 2001).

ويعتمد العائد على حالة الاقتصاد في وقت ما في المستقبل (غالبًا ما تسمى حالة الطبيعة)، والتي يمكن أن تكون في حالة نمو قوي، أو ركود، أو تضخم. وفيما يلي العوائد السنوية التي يقدرها الخبراء:

- إذا كان الاقتصاد ينمو بقوة، فإن عائد السندات سيكون ١٢٪، والأسهم ١٥٪، والودائع لأجل ٦,٥٪.
- وفي حالة أن يكون الركود سائدًا، فإن عائد السندات سيكون ٦٪، والأسهم ٣٪، والودائع لأجل ٦,٥٪.
- وإذا كان التضخم هو السائد، فإن عائد السندات سيكون ٣٪، أما الأسهم فستحقق خسارة ٢٪، وستحقق الودائع لأجل عائد ٦,٥٪.

وتكمن المشكلة في اختيار أفضل بديل للاستثمار. ومن المفترض أن تكون هذه البدائل منفصلة. وينبغي التعامل مع مجموعات مثل استثمار ٥٠٪ في السندات و٥٠٪ في الأسهم على اعتبار أنها بدائل جديدة.

ويمكن النظر إلى مشكلة اتخاذ القرار في الاستثمار على اعتبار أنها لعبة بين شخصين (انظر Kelly, 2002). فالمسـتثمر يصنع الاختيار (أي التحرك)، ومن ثَم تحـدث حالة طبيعية (يحدث التحرك). ويوضح الجدول (٦-٣) مردود نموذج رياضي؛ حيث يشتمل الجدول على متغيرات القرار (البدائل)، ومتغيرات لا يمكن السيطرة عليها (حالات الاقتصاد، مثل: البيئة)، ومتغيرات النتائج (العائد المتوقع، مثل: النتائج). وقد تم تصميم جميع النماذج في هذا القسم في إطار جدول بيانات.

وإذا كان هناك مشكلة في اتخاذ القرار في ظل التأكد، فإننا سنعرف ماذا سيكون الاقتصاد كما يمكننا بسـهولة اختيار أفضل استثمار. غير أن ذلك ليس هو الحال، ولذلك فإنه ينبغي علينا أخذ حالتَـي عدم التأكد والخطر في الاعتبار. فمع عـدم التأكد، فإننا لا نعرف احتمالات كل حالة من الطبيعة. ومع الخطر، سنفترض أننا نعرف الاحتمالات التي تحدث بها كل حالة من الطبيعة.

معالجـة عدم التأكد: هناك عدة طرق للتعامل مع عدم التأكد. فعلى سـبيل المثال، فإن النهج التفاؤلي يفترض حدوث أفضل نتيجة ممكنة لكل بديل ومن ثَم يختار أفضل الأفضل (أي: الأسـهم). أما النهج التشاؤمي فإنه يفترض حدوث أسوأ نتيجة ممكنة لكل بديل ثم يختار الأفضل من بين النتائج (أي: شهادات الإيداع CDs). وهناك نهج آخر يفترض ببساطة أن جميع حالات الطبيعة متساوية في الاحتـمالات (انظر Clemen & Reilly, 2000; Goodwin & Wright, 2000; Kontoghiorghes, Rustem, & Siokos, 2002). وعنـد التعامل مع حالـة عدم التأكد فإن كل نهج من هذه المناهج

لديه مشـاكل خطيرة عند تداوله لها. وينبغي على المحلل أن يحاول جمع ما يكفي من المعلومات قدر الإمكان، بحيث يمكن التعامل مع المشكلة في إطار التأكد أو الخطر المفترضة.

معالجة الخطر: تُعد طريقة اختيار البديل بأكبر قيمة متوقعة هي الطريقة الأكثر شيوعًا لحل مشـكلة تحليل الخطر. ولنفترض أن الخبراء يقدرون فرصة النمو القوي عند ٥٠٪، وفرصة الركود عند ٣٠٪، وفرصة التضخم عند ٢٠٪، فإن جدول القرار عندئذٍ سـتتم إعادة كتابته مع الاحتمالات المعروفـة (انظر الجدول ٦-٣). كما يتم حسـاب القيمة عن طريـق ضرب النتائج (أي: النواتج) حسب الاحتمالات الخاصة بها وإضافتها. فعلى سبيل المثال، الاستثمار في السندات يؤدي إلى عائد متوقع قدره ٨,٤ = (٠,٢) ٣ + (٠,٣) ٦ + (٠,٥) ١٢٪.

وفي بعـض الأحيـان فإن هذا النهج قد يكون إسـتراتيجية خطرة نظـرًا لأن فائدة كل نتيجة محتملة ربما تكون مختلفة عن القيمة. وحتى في حالة وجود فرصة ضئيلة لخسـارة كارثية، فإن القيمة المتوقعة قد تبدو معقولة، غير أن المستثمر قد يكون غير مستعد لتغطية الخسارة. فعلى سـبيل المثال، لنفترض أن مستشارًا ماليًا يقدم لك اسـتثمارًا «شبه مؤكد» بقيمة ١٠٠٠$ فإنه من الممكـن أن يضاعف ذلك أموالك في يوم واحد، وعندئذٍ سـيقول لك المستشـار: "حسـنًا، هناك احتمال ٠,٩٩٩٩ بأنك ستضاعف أموالك، ولكن للأسف هناك احتمال ٠,٠٠٠١ أنك ستكون مسؤولًا عن خسارة ٥٠٠,٠٠٠$ من حسابك". وتكون القيمة المتوقعة لهذا الاستثمار كالتالي:

= ($١,٠٠٠ - $٥٠٠,٠٠٠) ٠٠٠١+($١,٠٠٠ - $٢,٠٠٠) ٠,٩٩٩٩

= $٩٩٩,٩٠ - $٥٠,١٠

= $٩٤٩,٨٠

ومن الممكن أن تكون الخسـارة المحتملة كارثية لأي مستثمر صغير ليس مليارديرًا. واعتمادًا على قدرة المسـتثمر على تغطية الخسارة، فإن للاستثمار أدوات مختلفة متوقعة. ولابد أن تتذكر جيدًا أن المستثمر يتخذ القرار مرة واحدة فقط.

جدول ٦-٣: نموذج لجدول حالة الطبيعة (متغيرات لا يمكن السيطرة عليها)

البديل	النمو القوي (٪)	الركود (٪)	التضخم (٪)
السندات	١٢٠	٦,٠	٣,٠
الأسهم	١٥,٠	٣,٠	٢,٠-
شهادات الإيداع CDs	٦,٥	٦,٥	٦,٥

أشجار القرار:

تُعد شجرة القرار تمثيلًا بديلًا لجدول القرار (وللاطلاع على أمثلة، انظر Mind Tools Ltd., mindtools. com). وتقوم شجرة القرار بإظهار علاقات المشكلة بيانيًا كما يمكنها التعامل مع حالات معقدة في نموذج مضغوط. ومع ذلك، فإن شجرة القرارات قد تكون مُرهقة إذا كان هناك العديد من البدائل أو حالات الطبيعة. وتشتمل كل من TreeAge) Software Inc., treeage.com) TreeAge Pro (Palisade Corp., palisade. Com) وPrecisionTree على أنظمة تحليل شجرة القرار القوية والبديهية والمتطورة. كما يُقدم هؤلاء الموردون أيضًا أمثلة ممتازة لأشجار القرار المستخدمة في الممارسة. ومن المُلاحظ أن مصطلح شجرة القرار قد استُخدم لوصف نوعين مختلفين من النماذج والخوارزميات، في حين أنه في السياق الحالي تشير أشجار القرار إلى تحليل السيناريو. ومن جانب آخر، فإن بعض خوارزميات التصنيف في التحليل التنبؤي (انظر الفصلين ٤ و٥) أيضًا خوارزميات شجرة القرار.

وقد ورد في الجدول (٦-٤) حالة مبسطة للاستثمار لأهداف متعددة (وهي حالة اتخاذ قرار لتقييم البدائل بالعديد من الأهداف المتناقضة أحيانًا). وتتمثل الأهداف (المعايير) الثلاثة في كل من العائد والسلامة والسيولة. وهذه الحالة تم اتخاذ القرار فيها في ظل التأكد المفترض؛ بمعنى أنه يتم توقع نتيجة واحدة ممكنة لكل بديل؛ كما يمكن أخذ حالات الخطر أو عدم التأكد الأكثر تعقيدًا هي الأخرى في الاعتبار. وبعض النتائج تكون نوعية (مثل: منخفضة، عالية) بدلًا من أن تكون رقمية.

جدول ٦-٤: أهداف متعددة

البديل	العائد (٪)	السلامة	السيولة
السندات	٨,٤	مرتفعة	مرتفعة
الأسهم	٨,٠	منخفضة	مرتفعة
شهادات الإيداع CDs	٦,٥	مرتفعة جدًا	مرتفعة

ولمزيد من المعلومات حول تحليل القرار، انظر Clemen و Reilly (2000)، وGoodwin و (2000) Wright، بالإضافة إلى مجتمع تحليل القرار Decision Analysis Society (informs.org/Community/ DAS). وعلى الرغم من أن ذلك الأمر معقد للغاية، إلا أنه من الممكن تطبيق البرمجة الرياضية مباشرةً في حالات اتخاذ القرار تحت المجازفة. وسوف نناقش العديد من الطرق الأخرى لمعالجة الخطر في جزء لاحق من الكتاب، وهذا يتضمن المحاكاة، وعوامل التأكد، بالإضافة إلى المنطق الغير واضح.

أسئلة مراجعة على القسم ٦-٨:

١- ما هو جدول القرار؟

٢- ما هي شجرة القرار؟

٣- كيف يمكن استخدام شجرة القرار في صنع القرار؟

٤- قم بوصف المقصود بأن يكون لديك أهداف متعددة.

٦-٩ مقدمة في المحاكاة:

نقدم في هذا القســم والقسم التالي مجموعةً من التقنيات التي تُستخدم لدعم القرار. وعلى نطاق واسع، فإن هذه الأساليب جميعها تقع تحت مظلة المحاكاة. والمحاكاة تمثل الواقع. وتعتبر المحاكاة في أنظمة القرار، تقنية لإجراء تجارب (مثل: تحليلات "ماذا - لو") مع جهاز حاسب على نموذج من نظام الإدارة. وبالمعنى الحرفي للكلمة فإن المحاكاة تعتبر طريقة وصفية أكثر من كونها إرشــادية. كما أنه ليس هناك بحث تلقائي عن الحل الأمثل. وبدلًا من ذلك، فإن نموذج المحاكاة يصــف أو يتنبــأ بخصائص نظام معين في ظل ظروف مختلفة؛ حيث يمكــن اختيار أفضل البدائل المتعددة عندما يتم حساب قيم الخصائص. وللحصول على تقدير (وتغير) للتأثير الكلي لإجراءات معينــة، فإنه عادةً ما تقوم عملية المحاكاة بتكرار التجربة عدة مرات. وتعتبر محاكاة الحاســب مناسبة لمعظم الحالات، وعلى الرغم من ذلك فإن هناك بعض عمليات المحاكاة اليدوية المعروفة (مثل: قيام قسم شرطة المدينة بمحاكاة جدولة سيارتها الدورية مع عجلة لعبة الكرنفال).

وتنطوي بعض حالات اتخاذ القرار الحقيقية في العادة، على بعض العشوائية. وبسبب تعامل العديد من مواقف القرار مع حالات شــبه منظمــة أو غير منظمة، فإن الواقع معقد، إلى الدرجة التــي قد تجعل من الصعوبة تمثيله عن طريق التحســين أو نماذج أخرى، ولكن يمكن اســتخدام المحــاكاة لمعالجته في كثير من الأحيان. وتُعد المحاكاة واحدة من أكثر طرق دعم القرار شــيوعًا. وللاطــلاع على مثال، انظر الحالة العملية (٦-٦). كما توضح الحالة العملية (٦-٧) قيمة المحاكاة في بيئة لا يتوفر فيها الوقت الكافي لإجراء التجارب التحليلية.

الخصائص الرئيسة للمحاكاة:

عادةً ما تتضمن المحاكاة بناء نموذج للواقع من أجل المجال العملي. ومقارنةً بالنماذج التحليلية الأخــرى، فإن نماذج المحاكاة قد تعــاني من قلة عدد الافتراضات حول حالــة القرار. وبالإضافة إلى

ذلك، فإن المحاكاة تُعد تقنية لإجراء التجارب. ولذلك، فإنها تشتمل على اختبار قيم محددة للقرار أو المتغيرات التي لا يمكن السيطرة عليها في النموذج مع ملاحظة التأثير على متغيرات المخرجات.

وأخيرًا، فإن المحاكاة لا تستخدم عادةً إلا عندما تكون المشكلة معقدة للغاية بحيث لا يمكن معالجتها باستخدام تقنيات التحسين الرقمي. ويقصد بالتعقيد في هذه الحالة إما عدم إمكانية صياغة المشكلة من أجل التحسين (فعلى سبيل المثال: بسبب الافتراضات الغير تطبيقية)، أو أن الصيغة كبيرة جدًا، أو أن هناك الكثير من التفاعلات بين المتغيرات، أو أن المشكلة عشوائية في طبيعتها (أي: تعرض خطر أو عدم التأكد).

حالة عملية ٦-٧

محاكاة آثار تدخلات الالتهاب الكبدي B

على الرغم من قيام الولايات المتحدة باستثمارات كبيرة في مجال الرعاية الصحية، غير أن هناك بعض المشاكل التي تبدو وكأنها تتحدى الحل. فعلى سبيل المثال، هناك نسبة كبيرة من السكان الآسيويين في الولايات المتحدة أكثر عرضة من غيرهم للإصابة بفيرس (B). وإذا لم يتم علاج المرض بشكل فعال، فإن واحدًا من كل أربعة أشخاص مصابين بالعدوى إصابة مزمنة مهدد بالإصابة بسرطان الكبد أو تليف الكبد، فضلًا عن المشاكل الاجتماعية المرتبطة بالمرض (مثل العزل). وقد يكون التحكم في هذا المرض مكلفًا للغاية. وهناك عدد من المعايير للتحكم في ذلك المرض، وتتضمن الفحص، والتطعيم، وإجراءات العلاج. كما تتردد الحكومة في إنفاق الأموال على أي طريقة للتحكم في المرض إلا إذا كانت فعالة من حيث التكلفة، كما أنه ليس هناك دليل على تحسن صحة المصابين بهذا المرض. وعلى الرغم من أن جميع معايير التحكم ليست مثالية لجميع الحالات، غير أن أفضل طريقة أو مجموعة طرق مكافحة هذا المرض غير معروفة حتى الآن.

المنهجية / الحل:

قام فريق متعدد التخصصات والذي تألف من أولئك الذين لديهم خلفيات طبية وإدارية، وهندسية، بتطوير نموذج رياضي يستخدم أساليب أبحاث العمليات (OR) والتي حددت التركيبة الصحيحة من إجراءات التحكم التي يجب استخدامها لمكافحة التهاب الكبد الوبائي B في سكان آسيا وجزر المحيط الهادئ. وعادةً ما يتم استخدام التجارب السريرية في المجال الطبي من أجل تحديد أفضل مسار للعمل في علاج الأمراض والوقاية

منهـا. ويتمثل تعقيد هذه الحالة في كون الالتهاب الكبدي B يسـتغرق فترة طويلة غير معتادة من الوقت حتى يحدث تقدم. وقد تم استخدام نماذج وطرق OR بسبب التكلفة العالية التي تصاحب التجارب السريرية في هذه الحالة. وقد قدمت مجموعة من نماذج ماركـوف Markov ونماذج اتخـاذ القرار طريقة أكثر فعالية من حيـث التكلفة لتحديد تركيبة معايير الرقابة التي يجب اسـتخدامها في أي وقت. ويساعد نموذج القرار في قياس الفوائـد الاقتصادية والصحية لمختلف الإمكانات الخاصة بالفحص والعلاج والتنقيح. كما يُسـتخدم نموذج ماركوف Markov لنمذجة تطور الالتهاب الكبدي B. وقد تم إنشـاء النموذج الجديد اسـتنادًا على المؤلفات والخبرات السـابقة من أحد الباحثين بالإضافة إلى ما يُسـتمد من البيانات الفعلية الحالية للعدوى والعلاج. وقد قام صانعو السياسة ببناء النموذج الجديد باستخدام Microsoft Excel لأنه سهل الاستخدام.

النتائج / الفوائد:

وقـد تم تحليل النموذج الناتج مقارنةً ببرامج التحكم الموجودة في كل من الولايات المتحـدة والصـين. ففـي الولايات المتحـدة تم تطوير أربـع إسـتراتيجيات ومقارنتها بالإستراتيجية الحالية. وتتمثل هذه الإستراتيجيات الأربع في:

١- يتم تطعيم جميع الأفراد.

٢- يتـم فحص الأفراد أولًا لتحديد ما إذا كانوا مصابـين بعدوى مزمنة. فإذا كانت الإجابة بنعم، فعندئذ يتم التعامل معهم.

٣- يتم فحص الأفراد أولًا لتحديد ما إذا كان لديهم عدوى مزمنة. فإذا كان لديهم العـدوى، يتم علاجهم. وبالإضافة إلى ذلك، يتم أيضًا فحص المرضى المقرّبين للمصابين وتلقيحهم إذا لزم الأمر.

٤- يتـم فحـص الأفـراد أولًا لتحديد مـا إذا كانوا مصابـين بعدوى مزمنـة أو بحاجة إلى تطعيم. فإذا كانوا مصابين، يتم علاجهم. وإذا كانوا بحاجة إلى التطعيم، يتم تطعيمهم.

وقد أشـارت نتائج المحـاكاة إلى أن إجراء اختبارات الدم لتحديـد العدوى المزمنة وتطعيم الأشخاص المقربين من الناس المصابين تُعد فعالة من حيث التكلفة.

وفي الصين، سـاعد النموذج على تصميم سياسـة تطعيم للأطفال والمراهقين. وقد تمت مقارنة هذه السياسة مع مستويات التغطية الحالية للتطعيم ضد الالتهاب الكبدي B. وقد انتهت إلى أنه عندما يتم تطعيم الأفراد دون سـن ١٩عامًا، فإن ذلك سـيؤدي إلى تحسـين

النتائج الصحية على المدى الطويل. وفي الواقع، فإن هذه السياسة كانت فاعليتها من حيث التكلفة المالية أكثر من فاعلية السياسة الحالية لمكافحة المرض في مكان ما في وقت التقييم.

أسئلة للمناقشة:

١- وضح ما تتميز به طرق OR مثل المحاكاة على طرق التجارب السريرية في تحديد أفضل قياس للسيطرة على الالتهاب الكبدي B.

٢- ما هي الطرق التي يوفر بها القرار ونموذج ماركوف Markov وسائل فعالة من حيث التكلفة لمكافحة المرض؟

٣- ناقــش كيف تكــون الخلفية متعددة التخصصات أصلًا في إيجاد حل للمشــكلة الموضحة في القضية.

٤- إلى جانــب الرعاية الصحيــة، اذكر في أي مجال يمكن لمثل هذا النهج النمطي أن يساعد في تقليل التكلفة.

Source: Adapted from Hutton, D. W., Brandeau, M. L., & So, S. K. (2011). Doing good with good OR: Supporting cost-effective Hepatitis B interventions. Interfaces, 41(3), 289300-.

مزايا المحاكاة:

تُستخدم المحاكاة في نمذجة دعم القرار للأسباب التالية:

- تُعد هذه النظرية واضحة إلى حد ما.
- من الممكن تحقيق قدر كبير من ضغط الوقت، مما يمنح المدير الشعور سريعًا ببعض التأثيرات للعديد من السياسات على المدى الطويل (من١ إلى ١٠سنوات).
- تُعد المحاكاة وصفية وليست معيارية. وهذا يسمح للمدير بطرح أسئلة "ماذا - لو". كما يمكن للمدير اســتخدام نهج التجربة والخطأ لحل المشــاكل، كما يمكن أن يفعل ذلك بشــكل أسرع، وعلى حساب أقل، وبدقة أكبر، بأقل المخاطر.
- كــما يمكن للمــدرب أن يقوم بالتجريب من أجل تحديد متغيرات القــرار وأي أجزاء من البيئة مهمة بالفعل، ومع بدائل مختلفة.
- يتطلب نموذج المحاكاة الدقيقة معرفة وثيقة بالمشــكلة، مما يجبر مصمم النموذج على التفاعل باستمرار مع المدير. وهذا الأمر مطلوب من أجل تطوير أنظمة دعم القرار لأن كلًا من المطور والمدير يكتسبان فهمًا أفضل للمشكلة، وللقرارات المحتملة المتاحة.
- يتم تصميم النموذج من منظور المدير.

- يُصمم نموذج المحاكاة من أجل مشكلة معينة وغالبًا لا يمكن حل أي مشكلة أخرى. وبالتالي، لا يحتاج المدير إلى أي فهم عام؛ لأن كل مكون في النموذج يتوافق مع جزء من النظام الحقيقي.
- يمكن للمحاكاة التعامل مع مجموعة كبيرة جدًا من أنواع المشـاكل، مثل: المخزون، والموظفين، بالإضافة إلى الوظائف الإدارية عالية المستوى، مثل: التخطيط بعيد المدى.
- من الممكن أن تتضمن المحاكاة بشـكل عام التعقيدات الحقيقية للمشـاكل؛ فالتبسيط ليس ضروريًّا. وعلى سبيل المثال، قد تستخدم المحاكاة توزيعات احتمالية حقيقية بدلًا من توزيعات نظرية تقريبية.
- تنتج المحاكاة العديد من مقاييس الأداء المهمة تلقائيًّا.
- غالبًا ما تكون المحاكاة هي الأسـلوب الوحيد لنمذجة أنظمة دعم القرار والتي يمكنها التعامل بسهولة مع المشاكل غير المنظمة نسبيًّا.
- وتتوفر بعض حزم المحاكاة سهلة الاستخدام نسبيًّا (مثل: محاكاة مونت كارلو Monte Carlo). وهي بالإضافة إلى اشـتمالها عـلى حزم جداول البيانات (مثـل: @RISK)، فإنها تتضمن أيضًا التأثير على برمجيات الرسم البياني، وحزم جافا Java (وغيرها من تطورات الويب)، وسوف تتم مناقشة أنظمة المحاكاة التفاعلية المرئية في وقت قريب.

عيوب المحاكاة:

تتمثل العيوب الرئيسة للمحاكاة فيما يلي:

- لا يمكن ضمان الحل الأمثل، لكن بشكل عام توجد حلول جيدة نسبيًّا.
- قـد تكون عمليـة تصميم نموذج المحاكاة بطيئة ومكلفة، في حـين أن أنظمة النمذجة الحديثة أسهل في الاستخدام من أي وقت مضى.
- عادةً ما تكون الحلول والاسـتدلالات من دراسـة المحاكاة غير قابلة للتحويل إلى مشاكل أخرى نظرًا لأن النموذج يتناول عوامل مشكلة فريدة من نوعها.
- أحيانًا يكون من السهل شرح المحاكاة للمديرين في حين أن الطرق التحليلية غالبًا ما يتم تجاهلها.
- في بعض الأحيان تتطلب برامج المحاكاة مهارات خاصة بسبب تعقيد طريقة الحل الرسمي.

منهجية المحاكاة:

تنطـوي المحـاكاة عـلى وضـع نمـوذج لنظـام حقيقـي وإجـراء تجـارب متكـررة عليه. أما منهجية المحاكاة فإنها تتكون من الخطوات التالية، كما هو موضح في الشكل (٦-١٢):

١- تحديد المشــكلة: علينا دراســة وتصنيف المشــكلة الواقعية، ثم نحدد لماذا يُعد نهج المحاكاة مناسبًا. وهنا يتم التعامل مع حدود النظام والبيئة وغيرها من جوانب توضيح المشكلة.

٢- بناء نموذج المحاكاة: وتنطوي هذه الخطوة على تحديد المتغيرات وعلاقاتها فضلًا عن جمع البيانات. وغالبًا ما يتم وصف العملية باستخدام مخطط انسيابي، ومن ثم تتم كتابة برنامج حاسب.

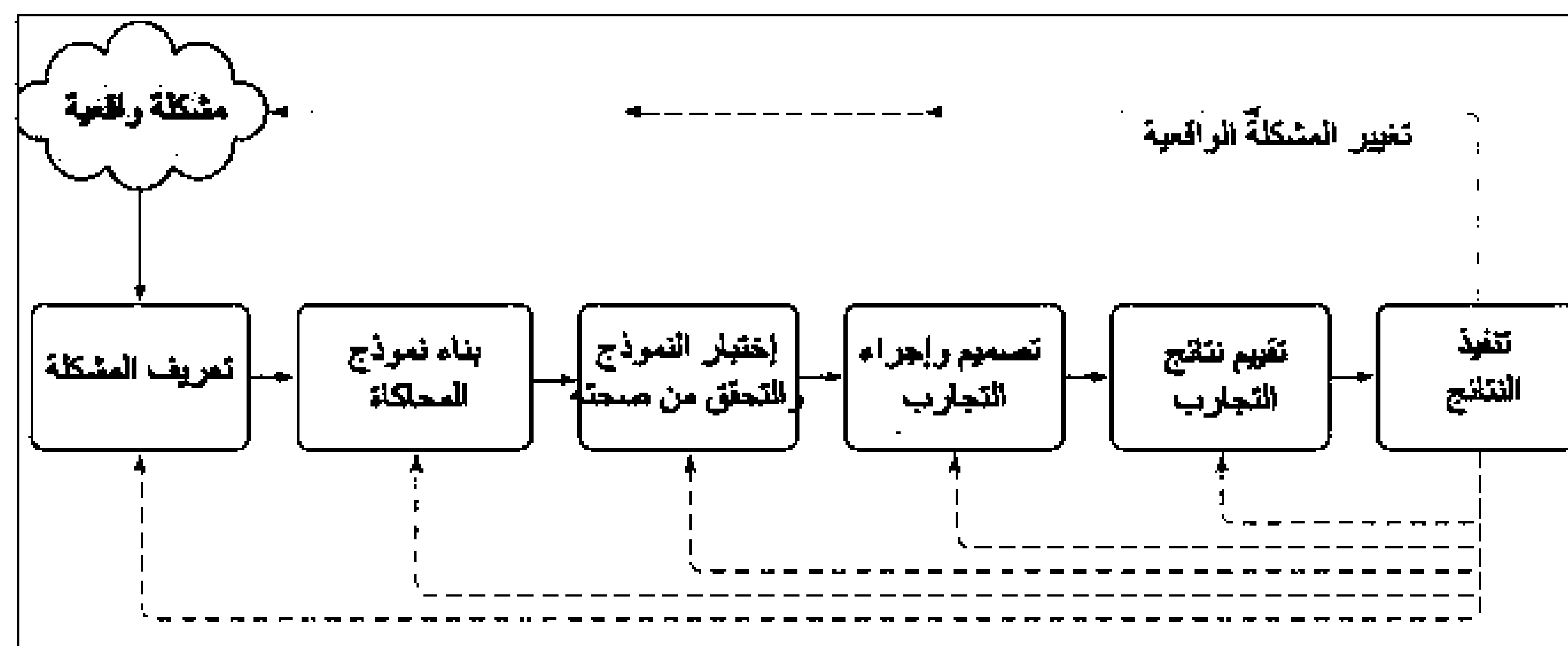
٣- اختبــار النموذج والتحقق من صحته: ينبغي أن يكــون نموذج المحاكاة ممثلًا للنظام الجاري دراسته بشكل صحيح، ولابد من اختبار ذلك والتحقق من صحته.

٤- تصميم التجربة: عند التأكد من صحة النموذج، يكون بذلك قد تم تصميم التجربة. ولابد من تحديد مدة تشغيل المحاكاة إذ إنها تُعد جزءًا من هذه الخطوة. وهناك نوعان من الأهداف المهمة والمتناقضة وهما: الدقة والتكلفة. ومن الحكمة أيضًا تحديد الحالات النموذجية (مثل: المتوسط والوسيط للمتغيرات العشوائية)، والحالات الأفضل (مثل: التكلفة المنخفضة، والربح المرتفــع)، بالإضافة إلى أســوأ حالة من الســيناريوهات (مثل: التكلفــة المرتفعة، والإيرادات المنخفضة)؛ حيث يســاعد ذلك في إنشاء نطاقات متغيرات القرار والبيئة التي تعمل فيها كما يساعد أيضًا في تصحيح نموذج المحاكاة.

٥- إجراء التجربة: ويقتضي إجراء التجربة مسائل تتراوح من إنشاء الأرقام العشوائية وصولًا إلى عرض النتائج.

٦- تقييم النتائج: ينبغي تفسير النتائج. بالإضافة إلى الأدوات الإحصائية القياسية، كما يمكن أيضًا استخدام تحليلات الحساسية.

٧- تنفيــذ النتائج: إن تنفيــذ المحاكاة مثله مثل أي تنفيذ آخر، فهو يتضمن نفس المشــاكل التي يتضمنهــا أي تنفيذ آخر. ومع ذلك، فإن فرص النجاح في المحاكاة تكون أفضل نظرًا لأن المدير عــادةً ما يكون مشــاركًا في عملية المحاكاة أكثر من النماذج الأخرى؛ حيث تؤدي المســتويات العليا من المشاركة الإدارية بشكل عام إلى مستويات أعلى من نجاح التنفيذ.



شكل ٦-١٢: عملية المحاكاة

وقد قدم كل من Banks و(2009) Gibson بعض النصائح المفيدة حول ممارسات المحاكاة؛ حيث قاموا على سبيل المثال، بسرد المشاكل السبعة التالية باعتبارها الأخطاء الشائعة التي يرتكبها مصممو نماذج المحاكاة. كما توفر القائمة، وإن لم تكن شاملة، اتجاهات عامة لمحترفي العمل بمشاريع المحاكاة.

- التركيز على النموذج أكثر من التركيز على المشكلة.
- توفير تقديرات للنقطة.
- عدم معرفة وقت التوقف.
- إعداد تقرير بما يريد العميل سماعه وليس بما تقوله نتائج النموذج.
- عدم فهم الإحصاءات.
- الخلط بين السبب والنتيجة.
- الفشل في مطابقة الواقع.

وفي مقال متابع، يتابع كل من Banks و(2009) Gibson بتقديم مبادئ توجيهية إضافية، ينبغي الرجوع إليها في هذا المقال:

analytics-magazine.org/spring-2009/205-software-solutions-the-abcs-of-simulationpractice.html.

أنواع المحاكاة:

وكما رأينا، فإنه عندما تكون الدراسات التجريبية وتجريب الأنظمة الحقيقية باهظة الثمن أو مستحيلة في بعض الأحيان، فإنه يتم استخدام المحاكاة والنمذجة؛ حيث تسمح لنا نماذج المحاكاة بالتحقق من مختلف السيناريوهات المثيرة للاهتمام قبل القيام بأي استثمار. وعمليات المحاكاة في الواقع، يتم فيها تعيين العمليات التي تحدث في العالم الحقيقي في نموذج المحاكاة، حيث يتكون النموذج من علاقات، وبالتالي معادلات تقدم جميعًا عمليات العالم الحقيقي. وتعتمد نتائج نموذج المحاكاة في هذه الحالة على مجموعة من المَعْلمات المعطاة للنموذج كمدخلات.

وهناك العديد من نماذج المحاكاة مثل: محاكاة مونت كارلو Monte Carlo، أو نموذج الأحداث المنفصلة، أو نموذج العوامل القائمة، أو نموذج ديناميات النظام. ويُعد مستوى التجريد في المشكلة هو أحد العوامل التي تحدد نوع تقنية المحاكاة. وعادةً ما تُستخدم نماذج الأحداث المنفصلة والنماذج المعتمدة على العوامل لمستويات التجريد الوسطى أو المنخفضة. وبالإضافة إلى

ذلــك، فإنه عادةً ما يؤخــذ في الاعتبار العناصر الفردية مثل الأشــخاص والأجزاء والمنتجات في نماذج المحاكاة، في حين أن ديناميات الأنظمة أكثر ملاءمة للتحليل الكلي.

وســوف نقدم في القسم التالي، الأنواع الرئيســة من المحاكاة، وهي: المحاكاة الاحتمالية، والمحاكاة المعتمدة على الوقت والمســتقلة عــن الزمن، بالإضافة إلى المحاكاة المرئيــة. كما أن هناك العديد من تقنيــات المحاكاة الأخرى مثــل نمذجة ديناميات النظام، والنماذج القائمة على الوكيل. وكما ذكرنا آنفًا، فإن الهدف هنا هو تمكينك من إدراك إمكانات بعض هذه التقنيات بدلًا من جعلك خبيرًا في استخدامها.

المحاكاة الاحتمالية: في المحاكاة الاحتمالية، يكون هناك واحد أو أكثر من المتغيرات المســتقلة (مثل: الطلب في مشكلة جرد) هي الاحتمالية. وهذه المتغيرات تتبع توزيعات احتمالية معينة، والتي قد تكون إما توزيعات منفصلة أو مستمرة:

- التوزيعات المنفصلة: وتنطوي على حالة مع عدد محدود من الأحداث (أو المتغيرات) والتي قد تأخذ عددًا محدودًا فقط من القيم.
- التوزيعــات المســتمرة: وهي حالات ذات عدد غير محدود مــن الأحداث المحتملة والتي تتبع وظائف الكثافة، مثل التوزيع الطبيعي.

ويعرض الجدول (٦-٥) كلا نوعي التوزيعات.

جدول ٦-٥: التوزيعات المنفصلة مقابل الاحتمالية المستمرة

الطلب اليومي	الاحتمالية المنفصلة	الاحتمالية المستمرة
٥	٠,١٠	يتــوزع الطلب اليومــي توزيعًــا طبيعيًا بمتوســط ٧ وانحراف معياري ١,٢.
٦	٠,١٥	
٧	٠,٣٠	
٨	٠,٢٥	
٩	٠,٢٠	

المحاكاة المعتمدة على الوقت في مقابل المحاكاة المســتقلة عن الوقت: يشــير مصطلح الزمن المســتقل إلى حالة لا يهم فيها معرفة وقت وقوع الحدث بالضبط. فعلى ســبيل المثال، قد يكون لدينــا علم أن الطلب على منتج معين عبارة عن ثلاث وحدات في اليوم، غير أننا لا نهتم متى يتم

طلب العنصر خلال اليوم. وفي بعض الحالات، قد لا يكون الوقت عاملًا في المحاكاة على الإطلاق، كــما هو الحال عند تصميم التحكــم في المصانع بطريقة ثابتة. ومع ذلك، فإنه في مشــاكل خط الانتظار التي تنطبق على التجارة الإلكترونية، من الضروري معرفة الوقت الدقيق للوصول (لمعرفة ما إذا كان العميل سيضطر إلى الانتظار). وهذا الموقف ينطبق على الوقت المحدد.

محاكاة مونت كارلو Monte Carlo:

عادةً ما يتم اســتخدام أحــد النوعين التاليين من المحاكاة الاحتمالية، في معظم مشــاكل القرارات التجارية. وتُعد محاكاة مونت كارلو هي طريقة المحاكاة الأكثر شــيوعًا لمشاكل قرارات العمل. وتبدأ هــذه الطريقة عادةً بتصميم نموذج لمشــكلة القرار دون الحاجة إلى أخذ عــدم التأكد في الاعتبار لأي متغــيرات، ثم ندرك أن هنــاك مَعْلمات أو متغيرات معينة تكون غير مؤكــدة أو تتبع توزيع احتمالية مفترض أو تقديري، ويعتمد هذا التقدير على تحليل البيانات الســابقة، وعندئذٍ نبدأ بتشــغيل تجارب أخذ العينات. وتتكون تجارب أخذ العينات الجارية من إنشاء قيم عشوائية من المَعْلمات غير المؤكدة ومن ثم قيم الحوسبة للمتغيرات التي تتأثر بمثل هذه المَعْلمات أو المتغيرات؛ حيث تصل هذه التجارب الخاصة بأخذ العينات بشــكل أســاسي إلى حل نفس النموذج مئات أو آلاف المرات. وبعد ذلك، يمكننا تحليل ســلوك هذه المتغيرات التابعة أو متغيرات الأداء من خلال دراسة التوزيعات الإحصائية الخاصة بهم. وقد اســتخدمت هذه الطريقة في محاكاة النظــم الفيزيائية وكذلك نظم الأعمال. ويوفر موقع Palisade.com (http://www.palisade.com/risk/monte_carlo_simulation.asp) برنامجًــا تعليميًا عامًا وجيدًا لطريقة محاكاة مونت كارلو. كما يقوم الموقع بتسويق أداة تسمى RISK@، وهي عبارة عن برنامج محاكاة شائع يعتمد على جداول بيانات مونت كارلو. كما يوجد برنامج آخر شائع في هذه الفئة وهو Crystal Ball، والذي يتم تسويقه حاليًا بواسطة Oracle as Oracle Crystal Ball. وبالطبع، فإنه يمكن أيضًا إنشــاء تجارب مونت كارلو وتشغيلها في جدول بيانات Excel دون استخدام أي برنامــج إضافي مثل البرنامجَين اللذَين ذُكِرا آنفًا. غير أن هذه الأدوات تجعله أكثر ملاءمة لتشــغيل مثــل هــذه التجارب في النماذج القائمة على Excel. الجديــر بالذكر أن نماذج محاكاة مونت كارلو قد تــم اســتخدامها في العديد من التطبيقــات التجارية. ومن الأمثلة على ذلك اســتخدام شركة بروكتور وجامبــل (Procter & Gamble) لهــذه النــماذج لتحديد مخاطر الصرف الأجنبي؛ واســتخدام شركة Lilly للنموذج لتحديد القدرة النباتية المثلى؛ وكذلك شركة أبو ظبي للمياه والكهرباء والتي استخدمت Risk@ للتنبؤ بالطلب على المياه في أبو ظبي؛ بالإضافة إلى الآلاف من دراســات حالات واقعية أخرى. ويشتمل كل موقع من مواقع شركات برامج المحاكاة على العديد من قصص النجاح.

محاكاة الحدث المنفصل:

تشـير محـاكاة الحدث المنفصـل إلى بناء نموذج لنظام يتم فيه دراسـة التفاعـل بين الكيانات المختلفة. وأبسـط مثال على ذلك هو مثال المتجر الذي يتألـف من خادم وعملاء. ويمكننا تقدير متوسـط أداء النظـام، ووقت الانتظار، وعدد العمـلاء المنتظرين، وما إلى ذلـك، من خلال نمذجة العمـلاء الذين يصلون إلى معـدلات مختلفة، والخادم الذي يخدم بمعـدلات مختلفة. ومثل هذه الأنظمـة يتـم النظر إليها باعتبارها مجموعـات من العملاء، وقوائم انتظـار، وخوادم. كما يوجد الآلاف مـن التطبيقات الموثقة لنماذج محاكاة الأحداث المنفصلة في الهندسـة، والأعمال التجارية، وما إلى ذلك. وقد كانت أدوات بناء نماذج محاكاة الحدث المنفصلة موجودة منذ فترة طويلة، غير أنها تطورت للاسـتفادة من التطورات في القدرات الرسـومية لبناء وفهـم نتائج مثل هذه النماذج للمحاكاة. وفي القسم التالي سوف نناقش هذه الطريقة الخاصة بالنمذجة. وتعرض الحالة العملية (٦-٨) مثالًا على اسـتخدام هذه المحاكاة في تحليل تعقيدات سلسلة التوريد التي تستخدم محاكاة مرئية والتي سيتم وصفها في القسم التالي.

حالة عملية ٦-٨

استخدام شركة كوزان (Cosan) المحاكاة من أجل تحسين سلسلة التوريد للطاقة

Cosan هـي مجموعـة شركات مقرها البرازيل وتعمل على مسـتوى العالم، وأحد أنشـطتها الرئيسـة هي زراعة ومعالجة قصب السـكر. وقد أصبح قصب السكر الآن مصدرًا رئيسًـا للإيثانول، فضلًا عن أنه المصدر الرئيس للسـكر، كما أنه مكون رئيس في الطاقة المتجددة. وقد أصبح الإيثانول نشاطًا كبيرًا لشركة Cosan، بسبب تزايد الطلب عـلى الطاقة المتجددة، حيث يقوم الإيثانول الآن بتشـغيل مصفاتـين بالإضافة إلى ١٨ مصنـع إنتاج، وملايـين الهكتارات hectares من مزارع قصب السـكر. ووفقًا للبيانات الحديثة فقد عالجت الشركة أكثر من ٤٤ مليون طن من قصب السكر، كما أنتجت أكثر من ١,٣ مليار لتر من الإيثانول، بالإضافة إلى ٣,٣ مليون طن من السكر. وكما يتبين، فإن عمليات هذا النطاق تؤدي إلى سلاسـل التوريد المعقدة. ولذلك؛ فقد طُلب من فريق الإمدادات تقديم توصيات إلى الإدارة العليا من أجل:

- تحديد العدد الأمثل من الشـاحنات المطلوبة في أسـطول يُسـتخدم لنقل قصب السكر إلى معامل التصنيع من أجل الحفاظ على رأس المال.

- تقديم اقتراح لكيفية زيادة قدرة قصب السكر الفعلية التي يتم إمداد مصانع السكر بها.

- تحديد مشـاكل عنق الزجاجة المتعلقة بالإنتاج لوضع حلول لها من أجل تحسين تدفق قصب السكر.

المنهجية / الحل:

قـام فريق الإمدادات بالعمل من خلال برمجيات Simio ومن ثَم قام الفريق ببناء نموذج محاكاة مركب لسلسـلة توريد شركة Cosan فيما يتعلق بهذه المشـاكل. وتبعًا للموجـز الذي قدمته برمجيات Simio، فإنه: "على مدار ثلاثة أشـهر، قام المهندسـون المستأجرون حديثًا بجمع البيانات المتعلقة بهذا المجال، كما تلقوا تدريبًا عمليًّا ومساعدة في النمذجة من Paragon Consulting of San Palo".

ولنمذجة العمليات الزراعية لتحليل رحلة ما بعد حصاد قصب السـكر حتى يصل إلى مصانع الإنتاج، فقد شملت أهداف النموذج تفاصيل الأسطول الناقل لقصب السكر عبر الطرق إلى Unity Costa Pinto، كما شملت أيضًا القدرة الفعلية لاستقبال مصانع سـكر القصب، والاختناقات، بالإضافة إلى نقاط التحسين في تدفق CCT (قطع التنزيل) من قصب السكر، وهكذا. وتتمثل مَعْلمات النموذج فيما يلي:

- متغيرات المدخلات: ٣٢.
- متغيرات الإنتاج: ٣٩.
- المتغيرات الإضافية: ٩٢.
- كيانات متغيرة: ٨.
- جداول المدخلات: ١٩.
- أيام المحاكاة: ٢٤٠ (الموسم الأول).
- عدد الكيانات: ١٢ (١٠ أنواع مركبة حصاد لنقل قصب السكر).

النتائج / الفوائد:

وقـد قدمت هذه التحليلات التي نتجت عن نمـاذج Simio فكرة جيدة عن خطر التشـغيل خلال فترة ٢٤٠ يومًا بسبب الشكوك المختلفة. وقد تمكنت الشركة من اتخاذ قـرارات أفضل وتوفير أكثر من ٥٠٠,٠٠٠$ مـن هذه الجهود الخاصة بالنمذجة وحدها فقط، من خلال تحليل الاختناقات المختلفة وطرق التخفيف من هذه السيناريوهات.

أسئلة للمناقشة:

١- ما نوع الاضطرابات في سلسلة التوريد التي يمكن أن تحدث في نقل قصب السكر من الحقل إلى مصانع الإنتاج لتطوير السكر والإيثانول؟

٢- ما هي الأنــواع المتقدمة من التخطيط والتنبؤ التي قد تكون مفيدة في تخفيف مثل هذا الاضطراب؟

ما الذي يمكن أن نتعلمه من هذه الحالة العملية؟

هذه القصة القصيرة للحالة العملية التي نحن بصددها توضح قيمة تطبيق المحاكاة على أحد المشــاكل والتي قد يصعب معها بناء نموذج التحسين. ومن خلال دمج نموذج محاكاة الحدث المنفرد والمحاكاة التفاعلية المرئية (VIS)، يمكن تصور تأثير الانقطاعات في سلسلة التوريد بسبب فشل الأسطول، والوقت غير المتوقع في هذا الوضع، وغير ذلك، ومن ثَم التوصل إلى التصحيحات المخطط لها.

Sources: Compiled from Wikipedia contributors, Cosan, Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Cosan&oldid=713298536 (accessed July 10, 2016); Agricultural Operations Simulation Case Study: Cosan, http://www.simio.com/case-studies/Cosan-agricultural-logistics-simulation-software-case-study/agricultural-simulation-software- case-study-video-cosan.php, (accessed July 2016); Cosan Case Study: Optimizing agricultural logistics operations, http://www.simio.com/case-studies/Cosan-agricultural-logistics-simulation-software-case-study/index.php, (accessed July 2016).

أسئلة مراجعة على القسم ٦-٩:

١- اذكر خصائص المحاكاة.

٢- اذكر مزايا وعيوب المحاكاة.

٣- اذكر مع الشرح خطوات منهجية المحاكاة.

٤- اذكر مع الشرح أنواع المحاكاة.

٦-١٠ المحاكاة التفاعلية المرئية:

فيما يلي سنقوم بفحص الطرق التي تُبين صورة صانع القرار أثناء اتخاذ القرار في العمل. فهو يعمل من خلال ســيناريوهات لشتى البدائل. حيث تتغلب هذه الطرق القوية على بعض أوجه القصور في الطرق التقليدية كما تساعد على بناء الثقة في الحل الذي تم التوصل إليه نظرًا لإمكانية تصور هذه الطرق بشكل مباشر.

عيوب المحاكاة التقليدية:

تُعـد المحاكاة طريقة راسـخة ومفيدة، بالإضافـة إلى أنها وصفية، وتعتمـد على الرياضيات لاكتساب نظرة ثاقبة في حالات اتخاذ القرار المعقدة. وعلى الرغم من كل ما ذُكر، إلا أن المحاكاة عادةً لا تسمح لصانعي القرار برؤية كيفية تطور حل مشكلة معقدة (مضغوطة)، بالإضافة إلى أن صانعي القرار لا يمكنهم التفاعل مع المحاكاة (والتي ستكون مفيدة لأغراض التدريب والتدريس). وبشـكل عام، فإن المحاكاة تشـير إلى النتائج الإحصائية في نهاية مجموعة من التجارب. وبالتالي، فإن صانعي القرار ليسـوا جزءًا مكملًا في تطوير المحاكاة والتجريب، ولا يمكن اسـتخدام خبراتهم وحكمهم بشكل مباشر. فإذا كانت نتائج المحاكاة لا تتطابق مع حدس أو حكم صانع القرار، فإنه من الممكن حدوث فجوة ثقة في النتائج.

المحاكاة التفاعلية المرئية:

إن المحاكاة التفاعلية المرئية (VIS)، والتي تُعرف أيضًا باسم النمذجة التفاعلية المرئية (VIM) وحل المشـاكل التفاعلي المرئي، هي طريقة محاكاة تتيـح لصانعي القرار رؤية ما يفعله النموذج وكيف يتفاعل مع القرارات المُتخذة حسـب ما يتم صنعها. وقد اسـتخدمت هذه التقنية بنجاح كبـير في تحليل العمليات في عدة مجالات مثل سلسـلة التوريد والرعايـة الصحية، حتى يتمكن المُسـتَخْدِم من توظيف معرفته لتحديد وتجربة إستراتيجيات القرارات المختلفة أثناء التفاعل مع النموذج. ومن الممكن حدوث التعلم المُعزز حول مشكلة وتأثير البدائل التي تم اختبارها.

ولعرض تأثير القرارات الإدارية المختلفة، فإن المحاكاة التفاعلية المرئية (VIS) تستخدم شاشات الرسـوم المتحركة للحاسـب؛ حيث تختلف هذه الشاشـات عن الرسـومات العاديـة في أنها تتيح للمستخدم ضبط عملية صنع القرار ورؤية نتائج التدخل. ويُعد النموذج المرئي رسمًا بيانيًّا يُستخدم باعتباره جزءًا لا يتجزأ من عملية صنع القرار أو حل المشاكل، وليس مجرد جهاز اتصال فقط، حيث يستجيب بعض الأشخاص للعروض الرسومية بشكل أفضل من غيرهم. كما يستطيع هذا النوع من التفاعل مساعدة المديرين في دراسة موقف اتخاذ القرار.

ويمكن للمحاكاة التفاعلية المرئية (VIS) أن توضح أنظمة ثابتة أو حركية؛ حيث تعرض النماذج الثابتـة صورة مرئية لنتيجة بديل واحد في كل مرة. أمـا النماذج الحركية فإنها تعرض الأنظمة التي تتطـور مـع مرور الوقت، ويتم توضيح التطور عن طريق الرسـوم المتحركـة. وقد ارتبطت أحدث تقنيـات المحاكاة المرئية بمفهوم الواقع الافتراضي، حيث يتم إنشـاء عالم اصطناعي لأغراض متعددة، من التدريب إلى الترفيه إلى عرض البيانات في مشـهد اصطناعي. فعلى سـبيل المثال، يقوم الجيش الأمريكي باستخدام أنظمة VIS حتى تتمكن القوات البرية من الإلمام بالتضاريس أو المدينة من أجل

توجيه نفسـها بسرعة كبيرة. كما يسـتخدم الطيارون VIS من أجل التعرف على الأهداف من خلال محاكاة عمليات الهجوم. وبالإضافة إلى ذلك، فإن برنامج VIS قد يتضمن أيضًا إحداثيات GIS.

النماذج التفاعلية المرئية وDSS:

اسـتُخدِمت النـماذج التفاعلية المرئيـة في أنظمة دعم القـرار في العديد من قـرارات إدارة العمليات. وتتكون الطريقة من تحضير (مثل تحضير مضخة مياه) نموذج تفاعلي مرئي لمصنع (أو شركة) بوضعه الحالي، ومن ثَم يعمل النموذج بسرعة على الحاسب، مما يسمح للمديرين بمراقبة احتمالات كيفية عمل المصنع في المستقبل.

وخـير مثال على النماذج التفاعلية المرئيـة (VIM) هو إدارة خط الانتظار (الطابور). وعادةً ما يقوم برنامج مثل أنظمة دعم القرار بحسـاب عدة مقاييس لأداء بدائل القرار المختلفة (مثل: وقـت الانتظـار في النظام). تتطلب مشـاكل الطابور اسـتخدام المحاكاة؛ حيـث يمكن للنماذج التفاعلية المرئية (VIM) عند تشـغيل المحاكاة عـرض حجم الطابور أثناء تغييره كما يمكنها أيضًا تقديم إجابات بيانية حول ما إذا كانت الأسئلة تتعلق بالتغيرات التي تحدث في متغيرات الإدخال. وتقدم الحالة العملية (٦-٩) مثالًا لمحاكاة مرئية اسـتُخدمت بغرض استكشـاف تطبيقات تقنية تحديد ترددات الراديو (RFID) في إعداد التصنيع.

كـما يمكن أيضًا دمج تقنيتَي النماذج التفاعلية المرئية والذكاء الاصطناعي؛ حيث يضيف دمج كلا التقنيتـين عدة قدرات تتراوح من القدرة على بنـاء الأنظمة بيانيًّا إلى التعلم عن ديناميكيات النظام. فهذه الأنظمة - لاسيما تلك التي تم تطويرها للعسكرية وصناعة ألعاب الفيديو - لديها أساليب تفكير تجعلها قادرة على التصرف بمستوى ذكاء عالٍ نسبيًّا في تفاعلاتها مع المستخدمين.

حالة عملية ٦-٩

تحسـين قرارات جدولـة محل العمل من خلال الترددات اللاسـلكية RFID: التقييم القائم على المحاكاة

يسعى مزود خدمات التصنيع الخاص بالمكونات البصرية والكهروميكانيكية المعقدة إلى اكتسـاب الكفاءة في قرار الجدولة الخاص بمحل عملـه نظرًا لمعاناة عمليات المتجر الحالية من بعض المشاكل:

- عدم وجود نظام للتسـجيل في حالة وصول عناصر عملية workin (WIP) فعليًّا إلى محطـات العمل أو عند مغادرتها محطات تشـغيل العمـل بالإضافة إلى مدة بقاء WIPs الفعلية في كل محطة عمل.

- عدم قدرة النظام الحالي على مراقبة أو متابعة حركة كل WIP في خط الإنتاج بشكل فوري.

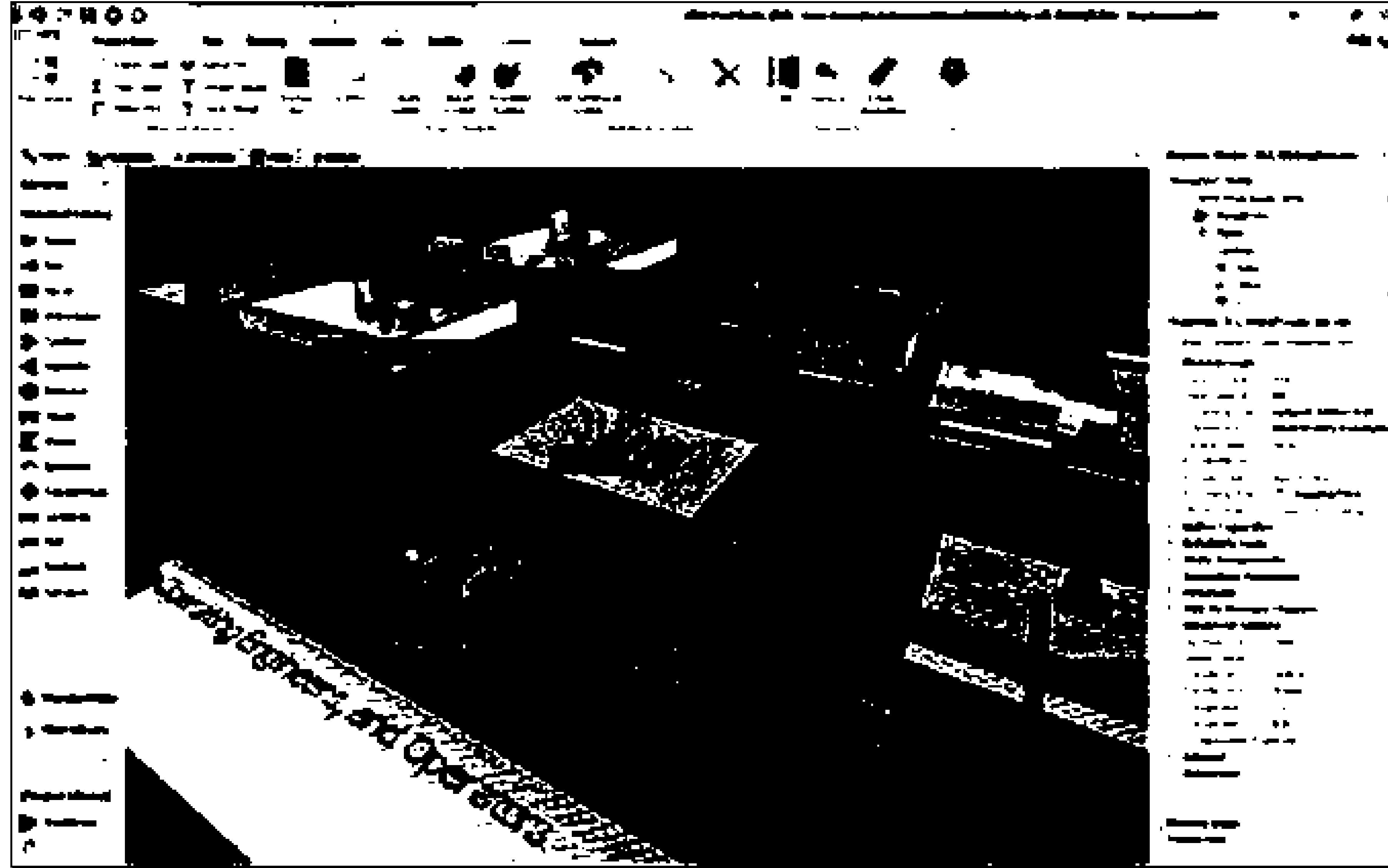
ونتيجة لذلك، فإن الشركة تواجه مشــكلتين رئيستين في هذا الخط الإنتاجي، وهي: مشــكلة التراكمات المرتفعة ومشــكلة تكاليف ارتفــاع العمل الإضــافي لتلبية الطلب. وبالإضافة إلى ذلك، فإن المنبع لا يمكنه الاســتجابة بسرعة كافية للحوادث غير المتوقعة مثل التغيرات في الطلب أو نقص المواد، كما أنه لا يمكنه مراجعة الجداول بطريقة فعالة مــن حيــث التكلفة. وتقوم الشركة بدراســة تنفيذ RFID على خــط الإنتاج. غير أنها بالرغم من ذلك، لا تعرف ما إذا كانت ستتحمل هذه النفقات الرئيسة من أجل إضافة رقائق RFID على صناديق الإنتاج، وتثبيت قارئات RFID في جميع أنحاء خط الإنتاج، وبالتالي بكل تأكيد ســتؤدي أنظمة معالجة هذه المعلومات إلى عدة مكاسب حقيقية. ولذلك؛ فإن هناك سؤالًا واحدًا وهو لاستكشاف أي التغيرات الجديدة في جدولة الإنتاج التي من المحتمل أن تَنتُج عن الاستثمار في بنية RFID التحتية.

المنهجية:

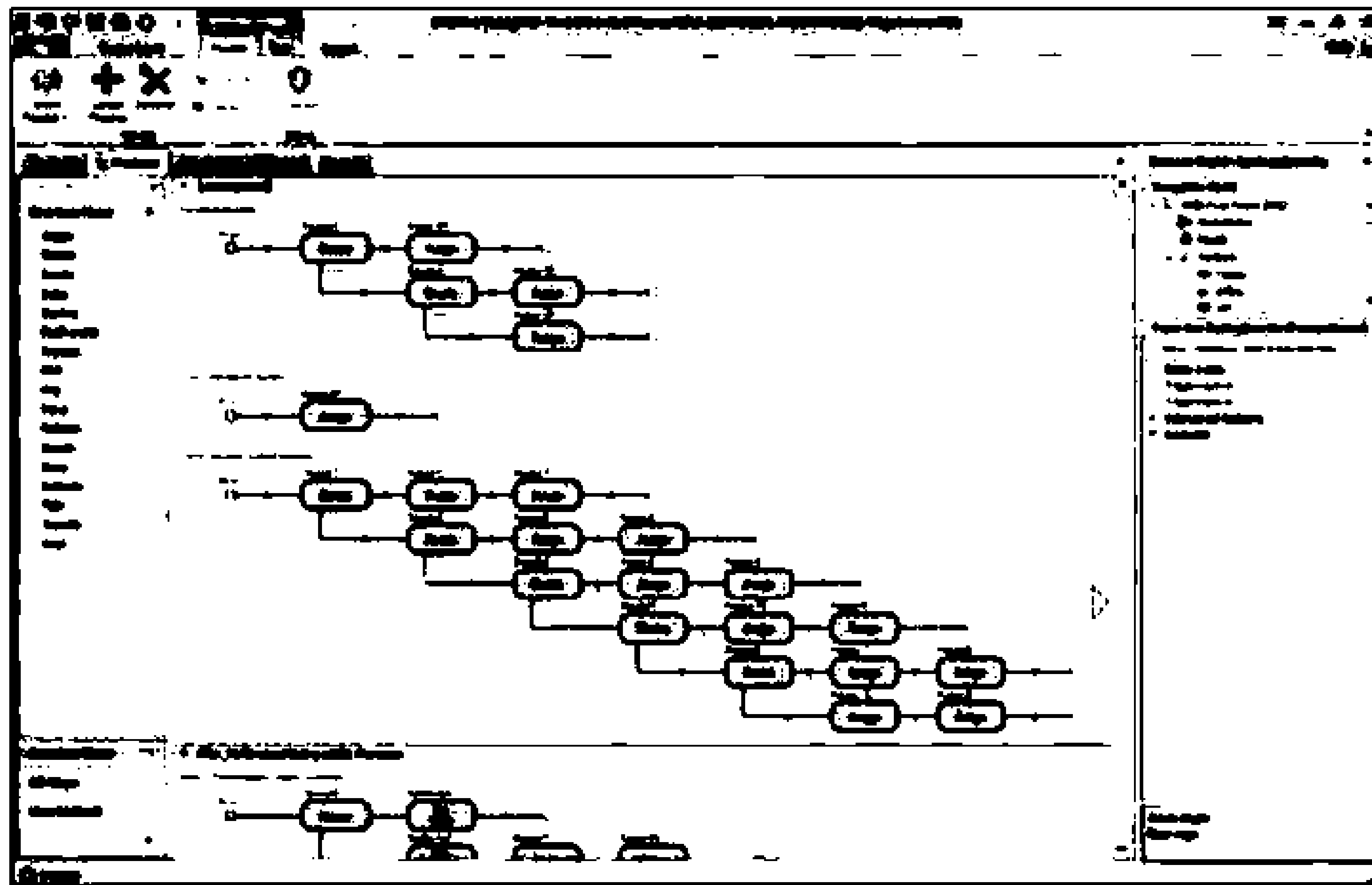
نظــرًا لأن استكشــاف إدخال أي نظام جديد في نظام الإنتــاج المادي من الممكن أن يكــون مكلفًــا للغاية أو حتى تخريبيًــا، فقد تم تطوير نموذج محــاكاة للحدث منفصل لفحص كيف يســتطيع التعقب (tracking) والتتبع من خلال RFID أن يقوم بتســهيل أنشطة جدولة إنتاج ورشة العمل. وقد تم اقتراح قاعدة جدولة VBS تعتمد على الرؤية وتستخدم أنظمة التتبع الفورية لتعقب أجزاء ومكونات WIPs، بالإضافة إلى المواد الخام في عمليــات الورش (shop-floor). وقــد تم تطبيق نهج المحاكاة من أجل فحص فائدة قاعدة VBS مقابل قواعد الجدولة الكلاســيكية: الأسبق في بدايات وأقدم قواعد موعد الاســتحقاق. وقد تم استخدام برامج SIMIO لتطوير نموذج المحاكاة. وبرامج SIMIO هي عبارة عن مجموعة من برامج محاكاة النماذج ثلاثية الأبعاد والتي تســتخدم أسلوبًا موجهًا نحو النمذجة، وقد تم استخدام هذه البرامج مؤخرًا في العديد من المجالات مثل المصانع وسلاسل التوريد والرعاية الصحية والمطارات وأنظمة الخدمات.

ويعرض الشكل (٦-١٣) لقطة شاشة من لوحة واجهة SIMIO لهذا الخط الإنتاجي. حيث تتضمن تقديرات المَعْلمات المُســتخدمة للحالة الأولية في نموذج المحاكاة كلًا من الطلبات والتوقعات الأسبوعية، بالإضافة إلى تدفق العملية، وعدد محطات العمل، وعدد مشــغلي الطوابق، ووقت التشــغيل في كل محطة عمل. وبالإضافــة إلى ذلك، فإنه يتم تقدير مَعْلَمات بعض بيانات الإدخال مثل وقت وضع علامات RFID أو وقت استرجاع المعلومــات، أو وقت تحديث النظام من خلال دراســة تجريبيــة ومن خبراء الموضوع. ويوضح الشــكل (٦-١٤) الرؤية العملية لنموذج المحاكاة حيث يتم تنفيذ أوامر محاكاة محــددة وترميزهــا. كما يقدم كلا الشــكلين (٦-١٥) و(٦-١٦) عرضًــا للتقرير القياسي

وتقرير الشـبكة المحورية لنموذج المحاكاة، حيث يوفر كل من التقرير القياسي وشـكل الشـبكة المحورية طريقةً سريعة جـدًا للعثور على نتائج إحصائية محددة، المتوسـط، أو النسـبة المئوية، أو الإجمالي، أو الحد الأقـصى، أو الحد الأدنى لقيم المتغيرات التي تم تعيينها والتقاطها كمخرج من نموذج المحاكاة.



شكل ٦-١٣: عرض واجهة SIMIO لنطام المحاكاة



شكل ٦-١٤: عرض عملية نموذج المحاكاة

شكل ٦-١٥: عرض تقرير قياسي

شكل ٦-١٦: تقرير شبكة محورية من تشغيل SIMIO

النتائج:

تشير نتائج المحاكاة إلى أن الأداء الذي يَنتُج عن قاعدة الجدولة المعتمدة على RFID يُعــد أفضل إذا ما قورِن بقواعد الجدولة التقليديــة فيما يتعلق بوقت المعالجة، ووقت

الإنتاج، واســتخدام الموارد، والتراكم، والإنتاجية. فإن الشركة يمكنها أخذ هذه المكاســب الإنتاجية ومن ثَم تقوم بتحليل التكلفة / الفائدة في اتخاذ قرارات الاستثمار النهائية.

أسئلة للمناقشة:

١- ما هي الأساليب الأخرى التي يمكن اتخاذها لتحليل قرارات الاستثمار في الحالات التي تشبه هذه الحالة؟

٢- كيــف يمكن توفير الوقت في حالة قدرة رقاقة RFID على معرفة الموقع الدقيق للمنتج قيد التشغيل؟

٣- ابحــث لتتعرف عــلى تطبيقات أجهزة استشــعار RFID في أماكــن أخرى، مع توضيح أي منها تجده أكثر إثارة للاهتمام.

Source: Based on Chongwatpol, J., & Sharda, R. (2013). RFID-enabled track and traceability in job-shop scheduling environment. European Journal of Operational Research, 227(3), 453- 463, http://dx.doi.org/10.1016/j.ejor.2013.01.009.

برامج المحاكاة:

هنــاك المئات من حزم المحاكاة التــي تتوفر لمجموعة من حالات اتخاذ القرار المتنوعة، حيث يتم تشــغيل الكثير منها كنظم على شبكة الإنترنت. وتقوم ORMS Today بنشر مراجعة دورية لبرامــج المحاكاة. وأحد هذه المراجعــات الحديثة (اعتبارًا من أكتوبــر ٢٠١٥) متاح عبر orms-today.org/surveys/Simulation/Simulation.html (تمت الزيارة في يوليو ٢٠١٦). وتتضمن حزم برامج الحاسب كلًا من:

- Analytica (Lumina Decision Systems, lumina.com).
- Crystal Ball Excel add-ins (والتــي تباع الآن من قِبل Oracle as Oracle Crystal Ball, oracle.com).
- @RISK (Palisade Corp., palisade.com).
- Arena (soldby Rockwell Intl., arenasimulation.com)، وهو برنامج تجاري كبير لمحاكاة الأحداث المنفصلة.
- Simio (simio.com)، وهــو برنامــج للمحاكاة التفاعلية المرئية (VIS) ســهل الاســتخدام تم تطويره من قِبَل مطوري برنامج Arena الأصليين. وبالإضافة إلى ذلك، فهناك:

- ExtendSim (extendsim.com)، وهــو برنامج محاكاة تفاعلية مرئية (VIS) آخر متميز كما أن له شهرة واسعة.
- JMP وتمتلكه SAS، وهو عبارة عن حزمة برمجية تحليلية رسومية تنطوي أيضًا على المحاكاة كأحد مكوناتها.

وللحصــول على معلومات حــول برامج المحاكاة، يمكنك مراجعة مجتمــع النمذجة والمحاكاة الدولية (scs.org) ومسح البرنامج السنوي في ORMS Today (ormstoday.com).

أسئلة مراجعة على القسم ٦-١٠:

١- عرف المحاكاة المرئية مع المقارنة بينها وبين المحاكاة العادية.

٢- قم بوصف مميــزات المحاكاة التفاعلية المرئية (VIS) (أي: النمذجة التفاعلية المرئية VIM) التي تجعلها جذابة لصناع القرار.

٣- كيف يمكن استخدام المحاكاة التفاعلية المرئية (VIS) في إدارة العمليات؟

٤- كيف يشبه فيلم الرسوم المتحركة تطبيق VIS؟

تلخيص لأهم نقاط الفصل:

- تلعــب النــماذج - متعددة الأنواع - دورًا رئيسًــا في أنظمة دعم القرار لأنها تُســتخدم لوصف مواقف حقيقية لصنع القرار.
- يمكن أن تكون النماذج ثابتة (أي لقطة مفردة لموقف ما) أو متحركة (أي فترة متعددة).
- يتــم إجراء التحليل تحت درجة تأكد مفترضة (والذي يعتبر مقبولًا بنســبة كبيرة) أو الخطر أو عدم التأكد (والذي يعتبر مقبولًا بنسبة ضئيلة).
- يعــرض المخططات البيانيــة التأثيرية بيانيًّا العلاقات الداخلية لنموذج ما. ويمكن اســتخدامها لتعزيز استخدام تقنية جداول البيانات.
- يوجــد لدى جداول البيانات العديد من القدرات، بمــا في ذلك تحليل "ماذا - لو"، والبحث عن الهدف، والبرمجة، وإدارة قواعد البيانات، والوصول للوضع الأمثل، والمحاكاة.
- يمكن لكل من جداول القرار وأشجار القرار أن تقوم بنمذجة وحل مشاكل بسيطة لصنع القرار.
- تعتبر البرمجة الرياضية منهجًا مهمًّا للوصول للوضع الأمثل.
- تعتبر البرمجة الخطية هي أكثر طرق البرمجة الرياضية انتشــارًا. فهي تحاول تحقيق التخصيص الأمثل للموارد المحدودة تحت قيود تنظيمية.
- تعتبر الأجزاء الرئيسة لنموذج البرمجة الخطية هي دالة الهدف، ومتغيرات القرار، والقيود.

- تعتبر حل مشاكل صنع القرار متعدد المعايير صعبًا ولكن ليس مستحيلًا.
- "ماذا - لو" و"البحث عن الهدف" هما أكثر الطرق انتشارًا لتحليل الحساسية.
- تشتمل العديد من أدوات تطوير أنظمة دعم القرار بداخلها على نماذج كمية (كنماذج مالية أو إحصائية) أو يمكنها التعامل بسهولة مع مثل هذه النماذج.
- تعتبر المحاكاة نهج أنظمة دعم القرار واسع الاستخدام؛ إذ تتضمن تجربة ذات نموذج يمثل موقفًا حقيقيًّا لصنع القرار.
- يمكن للمحاكاة التعامل مع حالات أكثر تعقيدًا من التحسين، ولكنها لا تضمن حلًا أمثل.
- توجد عدة طرق مختلفة للمحاكاة، ويعد بعضها مهمًّا في اتخاذ القرار ومنها محاكاة مونت كارلو ومحاكاة الحدث المنفصل.
- يسمح VIS/VIM لصانع القرار بالتفاعل المباشر مع أي نموذج ويعرض النتائج بأسلوب سهل الفهم.

مصطلحات أساسية:

التأكد.	البحث عن الهدف.	أهداف متعددة.	عدم التأكد.
تحليل القرار.	المخطط البياني التأثيري.	الحل الأمثل.	متغير غير قابل للتحكم فيه.
جدول القرار.	متغير النتيجة المتوسط.	مَعْلَمَة.	النمذجة التفاعلية المرئية (VIM).
شجرة القرار.	البرمجة الخطية (LP).	نموذج كمي.	المحاكاة التفاعلية المرئية (VIS).
متغير القرار.	البرمجة الرياضية.	متغير النتيجة (الحصيلة).	تحليل "ماذا - لو".
محاكاة الحدث المنفصل.	محاكاة مونت كارلو.	الخطر.	التوقع.
نماذج متحركة.	تحليل متعدد الأبعاد (نمذجة).	تحليل الخطر.	
المسح البيئي والتحليل.	النماذج الثابتة.	تحليل الحساسية.	المحاكاة.

أسئلة للمناقشة:

١- كيف ترتبط التحليلات التوجيهية بكل من التحليلات الوصفية والتنبؤية؟

٢- اشرح الاختلافات بين النماذج الثابتة والمتحركة. وكيف يمكن استنباط أحدهما من الآخر؟

٣- ما هو الفرق بين النهج التفاؤلي والنهج التشاؤمي في صنع القرار تحت درجة عدم التأكد المفترضة؟

٤- اشرح لماذا يتضمن حل المشـاكل تحت درجة عدم التأكد أحيانًا افتراض أن المشـكلة يجب أن يتم حلها تحت شروط الخطر.

٥- يحتمـل أن يكون Excel هو برنامج جداول البيانات الأكثر انتشـارًا بالنسـبة لأجهزة الحاسـب الشخصية، لماذا؟ وما الذي يمكننا القيام به مع هذه الحزمة لجعلها جذابة جدًا لجهود النمذجة؟

٦- اشرح كيفية عمل أشجار القرار. وكيف يمكن حل مشكلة معقدة باستخدام شجرة القرار؟

٧- اشرح كيف يمكن للبرمجة الخطية حل مشاكل التخصيص.

٨- ما هي مزايا وعيوب استخدام حزمة جداول البيانات لإنشاء وحل نماذج LP؟

٩- ما هي مزايا وعيوب استخدام حزمة البرمجة الخطية لإنشاء وحل نماذج LP؟

١٠- مـا هـو الفرق بـين تحليل القـرار مع هدف واحـد وتحليل القـرار مع أهـداف متعددة (أي المعايير)؟ اشرح الصعوبات التي قد تظهر عند تحليل الأهداف المتعددة.

١١- اشرح عمليًا كيف يمكن ظهور أهداف متعددة.

١٢- قارن بين تحليلي "ماذا - لو" و"البحث عن الهدف".

١٣- قم بوصف العملية العامة للمحاكاة.

١٤- اذكر بعض المزايا الرئيسة للمحاكاة فوق التحسين والعكس.

١٥- يمكن اعتبار العديد من ألعاب الحاسب محاكاة مرئية، اشرح لماذا.

١٦- اشرح لماذا تكون VIS مفيدة بشكل خاص في تنفيذ التوصيات المشتقة من أجهزة الحاسب.

تمارين - شبكة جامعة تيراداتا (TUN) وغيرها من التمارين اليدوية:

١- قم باستكشـاف teradatauniversitynetwork.com، وحدد كيف يتم استخدام النماذج في حالات ذكاء الأعمال والأوراق.

٢- قم بإنشاء نماذج جداول البيانات المعروضة في الشكلين (٦-٣) و(٦-٤).

١- ما هو تأثير تغير معدل الفائدة من ٨٪ إلى ١٠٪ في نموذج جدول البيانات المعروض في الشكل (٦-٣)؟

٢- بالنسبة للنموذج الأصلي في الشكل (٦-٣)، ما هو معدل الفائدة المطلوب لتخفيض الدفعات الشهرية بنسبة ٢٠٪؟ وما هو التغير في مبلغ القرض الذي سيكون له نفس التأثير؟

٣- في جدول البيانات المعروض في الشكل (٦-٤)، ما هو تأثير الدفع المقدم بمبلغ ٢٠٠ دولار كل شهر؟ ما هو الدفع المقدم الضروري لسداد القرض في ٢٥ عامًا بدلًا من ٣٠ عامًا؟

٣- قم بحل مشكلة مزيج المنتجات (MBI) التي تم وصفها في هذا الفصل، باستخدام Excel's Solver أو نسخة الطالب من برنامج LP solver، مثل Lindo. حيث يتاح Lindo من Lindo Systems,. Inc، على موقع lindo.com، وكذلك يتوفر غيره - ابحث في الويب. قم باختبار تقارير الحل (المخرجات) للإجابات وتقرير الحساسية. هل حصلت على نفس النتائج المذكورة في هذا الفصل؟ جرب تحليل الحساسية المشار إليه في الفصل وهو تخفيض الجانب الأيمن من القيود التسويقية CC-8 بمقدار وحدة واحدة، أي من ٢٠٠ إلى ١٩٩. ماذا يحدث للحل عندما تقوم بحل المشكلة بعد هذا التعديل؟ قم بإزالة قيد CC-8 الأدنى تمامًا (يمكن عمل ذلك بسهولة إما بحذفه في Solver أو جعل الحد الأدنى صفرًا) ثم قم بإعادة حل المشكلة. ماذا يحدث؟ باستخدام الصيغة الأصلية، حاول تعديل معاملات دالة الهدف وانظر ماذا يحدث.

٤- عبر البحث في الويب قم بفحص كيفية استخدام النماذج وحلولها بواسطة إدارة الأمن الوطني الأمريكي في "الحرب ضد الإرهاب". قم أيضًا بفحص كيفية استخدام الحكومات أو الهيئات الحكومية الأخرى للنماذج في مهامها.

٥- هذه المشكلة مساهمة من د. Rick Wilson من جامعة أوكلاهوما الحكومية.

لقد ضرب الجفاف مؤخرًا المزارعين بشدة. تأكل الأبقار الذرة الحلوة.

أنت مهتم بإعداد خطة تغذية للأسبوع القادم لمواشيك باستخدام سبعة منتجات غذائية غير تقليدية وهي: (١) حبوب شوكولاتة Lucky Charms، (٢) أصابع الزبدة، (٣) حليب، (٤) أيس كريم فانيليا، (٥) حبوب Cap'n Crunch، (٦) الذرة الحلوة (لأن الذرة الحقيقية كلها ميتة)، (٧) شرائح البسكويت Ahoy.

شرائح Ahoy	الذرة الحلوة	Cap'n Crunch	أيس كريم فانيليا	حليب	أصابع الزبدة	شوكولاتة Lucky Charms	
٦,٧٥	٤	٥,٢٥	٦,٣٥	٤,٢٥	٧	٢,١٥	دولار / رطل
نعم	لا	لا	لا	نعم	نعم	نعم	شوكولاتة
٦٢	٢٦	٧٢	٦٥	٤٥	٨٠	٧٥	بروتين
١٢	٨	١١	٦	١٨	٢٠	١٢	TDN

شرائح Ahoy	الذرة الحلوة	Cap'n Crunch	أيس كريم فانيليا	حليب	أصابع الزبدة	شوكولاتة Lucky Charms	
٥	١	٢	١٢	٤,٥	٤	٣	كالسيوم

وتظهر تكلفة الرطل بالدولار، وكذلك وحدات البروتين لكل رطل تساهم به، ومجموع العناصر الغذائية المنهضمة (TDN) التي يساهمون بها لكل رطل، ووحدات الكالسيوم لكل رطل.

أنت تقدر أن الكمية الإجمالية من منتجات التغذية غير التقليدية تساهم بالمقادير التالية من العناصر الغذائية: ٢٠٠٠٠ وحدة بروتين على الأقل، و٤٠٢٥ وحدة TDN على الأقل، ومن ١٠٠٠ على الأقل إلى ١٢٠٠ على الأكثر وحدة كالسيوم.

ويوجد أيضًا بعض المتطلبات المتنوعة الأخرى:

- لا يمكن أن تتجاوز الشوكولاتة في خطة التغذية الشاملة (بالرطل) مقدار العناصر الأخرى غير الشوكولاتة. ويعرض الجدول السابق ما إذا كان المنتج يعتبر شوكولاتة أم لا (نعم = شوكولاتة، لا = ليس شوكولاتة).
- لا يمكن أن يشكل أي منتج تغذية واحدًا أكثر من ٢٥٪ من الوزن الإجمالي (بالرطل) اللازم لإنشاء مزيج تغذية مقبول.
- هناك نوعان من الحبوب (شوكولاتة Lucky Charms وCap'n Crunch). لا يمكنهما مجتمعَين أن يكونا أكثر من ٤٠٪ (بالرطل) من المزيج الإجمالي لتلبية متطلبات المزيج.

حدد المستويات المثلى للمنتجات السبعة لوضع خطة تغذية أسبوعية خاصة بك بحيث تتدنَّى التكلفة. لاحظ أن كل مقادير المنتجات يجب ألا تحتوي على قيم كسرية (عدد الرطل أرقام صحيحة فقط).

٦- هذا التمرين مساهمة من د. Rick Wilson من جامعة أوكلاهوما الحكومية لتوضيح قدرات النمذجة لبرنامج Excel Solver.

أنت تعمل مع مجموعة كبيرة من العمال المؤقتين (مجموعة من المتدربين والمتقاعدين، إلخ) لوضع مسودة خطة لفريق عمل مركز اتصال ليلي (للمستقبل القريب). لديك أيضًا مجموعة صغيرة من العمال المتفرغين الذين هم "ثوابتك" - ولكنك وضعتهم بالفعل في الجدول، وقد أدى ذلك إلى متطلبات لتشغيل عمالة لديك. وأنت لا تهتم في النموذج بالعمال المتفرغين.

وتكون متطلبات تشغيل العمالة يومية وهي: ١٥ ليوم الاثنين، ٢٠ ليوم الثلاثاء، ١٩ ليوم الأربعاء، ٢٢ ليوم الخميس، ٧ ليوم الجمعة، ٣٢ ليوم السبت و٣٥ ليوم الأحد.

لديك من ٨ إلى ١٠ منهم لا يستطيعون العمل في عطلة نهاية الأسبوع (السبت أو الأحد).

بالنســبة لهؤلاء الناس "في يوم واحد فقط من أيام الأسبوع"، هناك ثلاث ورديات ممكنة: فسيعملون ٤ أيام من إجمالي ٥ أيام عمل في الأسبوع، وتحصل إحدى الورديات على عطلة في يوم الثلاثاء، ووردية أخرى في يوم الأربعاء، ووردية ثالثة في يوم الخميس.

وعليك أن تخصص ثمانية أشخاص على الأقل لهذه الورديات "في يوم واحد فقط من أيام الأسبوع".

وبالنســبة لجميع الورديات الأخرى (ولا تتقيد بحجم العمالة المشغلة)، يعمل الشخص ٤ أيام من إجمالي ٧ أيام كل أســبوع. وســوف يعمل العمال لمدة أسبوعين ويومَي عطلة نهاية الأســبوع (وردية "٢ / ٢"). وتعتبر جميع التوليفات «ذات اليومين» اليومية الممكنة ورديات عمل مرتبطة ماعدا أي توليفات يحصل فيها العمال على عطلة ثلاثة أيام متتالية، فتلك الأيام غير مسموح بها ولا يجب أن تكون في النموذج.

نحن نعمل بنموذج بســيط جدًا - وبدون تكاليف؛ إذ يكون الهدف من نموذجنا هو إيجاد أقل عدد من العمال الذي يلبي الحد الأدنى اليومي المطلوب لمركز الاتصال ولا يتم الاستعانة بأكثر من أربعة عمال إضافيين (فوق الحد الأدنى للمتطلبات) خلال أي يوم واحد.

كذلك، لا يمكن لجميع الورديات ("يوم واحد فقط من أيام الأســبوع" أو ورديات ٢ / ٢) أن «يخصص» لها أكثر من ستة أشخاص.

قم بإنشاء نموذج أساسي يقبل بهذه القيود ويعمل على تدنية العدد الإجمالي للأشخاص المطلوبين لتحقيق الحد الأدنى من المتطلبات. إذا كان هناك مشكلة، نعم، عدد الأشخاص عدد صحيح.

٧- هذا التمرين أيضًا مساهمة من د. Rick Wilson من جامعة أوكلاهوما الحكومية. يحاكي السيناريو البسيط التالي "Black Book" الموصوف في مقالة Business Week عن إنتاج الكوكاكولا من عصير البرتقــال. (-http://www.businessweek.com/articles/2013 -01- 31 coke-engineers-its (orange-juice-with-an-algorithm, accessed February 2013 قــم بإنشــاء نمــوذج برمجة خطية (LP) مناسب لهذا السيناريو.

بالنســبة للفترة الإنتاجية التالية، يوجد خمس دفعات مختلفة من عصير البرتقال الخام يمكن خلطها معًا لصنع منتجات عصير البرتقال، SunnyQ، GlowMorn، وOrenthal James.

في تكوين الخليط الأمثل من المنتجات الثلاثة من الدفعات الخمس المختلفة، يجب أن يهدف نموذج البرمجة الخطية إلى تعظيم صافي ســعر البيع لــكل جالون من المنتجات وتخفيض التكلفة المقدرة لكل جالون من العصير الخام.

ويتم هنا وصف خمس دفعات خام من عصير البرتقال من حيث درجة الحلاوة (Brix) واللب (Pulp) والمخزون المتاح (جالون) والتكلفة (دولار / جالون)، وكلها ذاتية التفسير:

- دفعة ١- أناناس برتقال A، ١٦ = brix، ١,٢ = pulp، ٢٥٠ جالون، ١,٢ دولار / جالون.
- دفعة ٢- أناناس برتقال B، ١٧ = brix، ٠,٩ = pulp، ٢٠٠ جالون، ٣٢,٢ دولار / جالون.
- دفعة ٣- متوسط الحلاوة، ٢٠ = brix، ٠,٨ = pulp، ١٧٥ جالون، ١٤,٣ دولار / جالون.
- دفعة ٤- فالنسيا، ١٨ = brix، ٢,١ = pulp، ٣٠٠ جالون، ٤١,٢ دولار / جالون.
- دفعة ٥- برتقال تيمبل، ١٤ = brix، ١,٦ = pulp، ٢٦٥ جالون، ٥٥,٢ دولار / جالون.

لاحظ أنه للتأكد من أن العصير الخام لا يصبح "قديمًا" جدًا بمرور الوقت، فإن أحد متطلبات الإنتاج هو أنه يتم بالضرورة استخدام ٥٠٪ على الأقل من المخزون المتاح من كل دفعة في خلط منتجات عصير البرتقال الثلاثة (وبالطبع لا يمكن استخدام أكثر من المتاح).

من منظور المنتج، يجب أن يكون هناك ١٠٠ جالون على الأقل من خليط SunnyQ و١٢٥ جالون على الأقل من كل من GlowMorn وOrenthalJames. وبالمثل، يشير الطلب المستقبلي المتوقع على المنتجات إلى أنه في هذه الفترة، يلزم وجود إنتاج تام بمقدار ٤٠٠ جالون على الأكثر من SunnyQ، و٣٧٥ جالونًا على الأكثر من GlowMorn، و٣٠٠ جالون على الأكثر من OrenthalJames.

لاحظ أيضًا، أنه عند خلط المنتجات من الدفعات الخمس، فإن الدفعة المنفردة لا يمكنها أن توفر أكثر من ٤٠٪ من إجمالي كمية أي منتج تام، ويتم تطبيق ذلك على كل منتج على حدة.

وتشمل سمات المنتجات الثلاثة ما يلي:

- سعر البيع.
- الحد الأقصى لمتوسط درجة حلاوة (brix) المنتج المخلوط النهائي.
- الحد الأدنى لمتوسط درجة حلاوة (brix) المنتج المخلوط النهائي.
- الحد الأقصى لمحتوى اللب.

وتشير متطلبات "المتوسط" الثلاثة إلى أن المتوسط المرجح لكل العصير المخلوط ببعضه لذلك يجب أن يحقق المنتج تلك المواصفات.

- SunnyQ: سعر البيع = ٩٢,٣ $ / جالون، ١٩ = Max Brix، ١٨,٥ = Min Brix، Max Pulp =١,٦.
- GlowMorn: سعر البيع = ١٤,٣ $ / جالون، ١٧ = Max Brix، ١٦,٧٥= Min Brix، Max Pulp = ١,٨.
- OrenthalJames: سعر البيع = ٧٧,٣ $ / جالون، ١٧,٧٥ = Max Brix، ١٧,٥٥ = Min Brix، Max Pulp = ١,١.

المراجع:

- Arsham, H. (2006a). Modeling and simulation resources. home.ubalt.edu/ntsbarsh/Business-stat/RefSim.htm (accessed July 2016).
- Arsham, H. (2006b). Decision science resources.home.ubalt.edu/ntsbarsh/Business-stat/Refop.htm (accessed July 2016).
- Bailey, M. J., Snapp, J., Yetur, S., Stonebraker, J. S., Edwards, S. A., Davis, A., & Cox, R. (2011). Practice summaries: American Airlines uses should-cost modeling to assess the uncertainty of bids for its full-truckload shipment routes. Interfaces, 41(2), 194-196.
- Banks, J., & Gibson, R. R. (2009). Seven sins of simulation practice. INFORMS Analytics, 24-27. www.analytics-magazine. org/summer-2009/193-strategic-problems-modeling-the-market-space (accessed July 2016).
- Neonatal physician scheduling at the University of Tennessee Medical Center. Interfaces, 46(2), 168-182.
- Businessweek.com. Coke engineers its orange juice—With analgorithm. www.businessweek.com/articles/2013-01-31/coke-engineers-its-orange-juice-with-an-algorithm (accessed July 2016).
- Chongwatpol, J., & Sharda, R. (2013). RFID-enabled track and traceability in job-shop scheduling environment. European Journal of Operational Research, 227(3), 453-463, http://dx.doi.org/10.1016/j.ejor.2013.01.009.
- Christiansen, M., Fagerholt, K., Hasle, G., Minsaas, A., & Nygreen, B. (2009, April). Maritime transport optimization: An ocean of opportunities. OR/MS Today, 36(2), 26-31.
- Clemen, R. T., & Reilly, T. (2000). Making hard decisions with Decision Tools Suite. Belmont, MA: Duxbury Press. Farasyn, I., Perkoz, K., & Van de Velde, W. (2008, July/August). Spreadsheet models for inventory target setting at Procter & Gamble. Interfaces, 38(4), 241-250.
- Furman, K. C., Song, J. H., Kocis, G. R., McDonald, M. K., & Warrick, P. H. (2011). Feedstock routing in the ExxonMobil downstream sector. Interfaces, 41(2), 149-163.
- Goodwin, P., & Wright, G. (2000). Decision analysis for management judgment, 2nd ed. New York: Wiley.
- Hurley, W. J., & Balez, M. (2008, July/August). A spreadsheet implementation of an ammunition requirements planning model for the Canadian Army. Interfaces, 38(4), 271-280.
- Hutton, D. W., Brandeau, M. L., & So, S. K. (2011). Doing good with good OR: Supporting cost-effective Hepatitis B interventions. Interfaces, 41(3), 289-300.
- Kearns, G. S. (2004, January-March). A multi-objective, multi-criteria approach for evaluating تقنية المعلومات investments: Results from two case studies. Information Resources Management Journal, 17(1), 37-62.
- Kelly, A. (2002). Decision making using game theory: An introduction for managers. Cambridge, UK: Cambridge University Press.
- Knight, F. H. (1933). Risk, uncertainty and profit: with an additional introductory essay hither to unpublished. London school of economics and political science.
- Koksalan, M., & Zionts, S. (Eds.). (2001). Multiple criteria decision making in the new millennium. Berlin: Springer-Verlag.

الفصل السابع
مفاهيم وأدوات البيانات الضخمة

أهداف التعلم:

- تعلَّم ما هي البيانات الضخمة وكيف تجعل عالم التحليلات يتغير.
- فهم الحافز لتحليلات البيانات الضخمة وقائدي العمل لها.
- الدراية التامة بنطاق عريض من تمكين التقنيات لتحليلات البيانات الضخمة.
- تعلَّم Hadoop وMapReduce وNoSQL من حيث ارتباطها بتحليلات البيانات الضخمة.
- مقارنة وتناقض الاستخدامات المكملة لمستودعات البيانات وتقنيات البيانات الضخمة.
- الدراية باختيار منصات وخدمات البيانات الضخمة.
- فهم الحاجة إلى تحليلات التيار (تدفق البيانات) وتقدير إمكانياتها.
- تعلَّم تطبيقات تحليلات التيار (تدفق البيانات).

لا تعتبر البيانات الضخمة - والتي تعني أشـياء كثيرة لكثير من الناس - تقليعة تقنية جديدة. فقد أصبحت من أولويات الأعمال التي من المحتمل أن تغير خريطة المنافسـة بشـكل عميق في الاقتصاد المتكامل على الصعيد العالمي في هذه الأيام، بالإضافة إلى توفير حلول ابتكارية لتحديات الأعـمال الدائمة؛ إذ تسـعى البيانات الضخمة والتحليلات لإيجاد طـرق جديدة لتحويل عمليات ومنظـمات وصناعات بأسْرِها وحتى المجتمع بالكامل. غير أن التغطية الإعلامية الشـاملة تجعل مـن الصعب التمييز بين الدعاية والواقع. ويهدف هذا الفصل إلى توفير تغطية شـاملة للبيانات الضخمـة وتقنياتها التمكينية ومفاهيم التحليلات المرتبطة بها؛ وذلك للمسـاعدة في فهم قدرات وقيود هذه التقنية حديثة الظهور. فيبدأ الفصل بتعريف البيانات الضخمة ومفاهيمها المرتبطة، يـلي ذلك التفاصيل الفنية للتقنيات التمكينية، والتي منها Hadoop وMapReduce وNoSQL. ونقـدم أيضًا تحليـلًا مقارنًا بين مسـتودعات البيانات وتحليلات البيانـات الضخمة، في حين يتم تخصيـص الجزء الأخير مـن الفصل لتحليلات التيار، والتي تعتبر إحـدى أقيم المقترحات الواعدة لتحليلات البيانات الضخمة.

٧-١ مقال افتتاحي: تحليل حركة العملاء في شـــركة اتصالات باستخدام طرق البيانات الضخمة:

خلفية:

أرادت شركة اتصالات (تسمى Access Telecom {AT} لأسباب تتعلق بالخصوصية) الحد من تحول العمـــلاء عن خدمـــات الاتصالات الخاصة بها. وبالرغـــم من أن انتقال العملاء بـــين الشركات في صناعة الاتصالات يعد أمرًا شـــائعًا، إلا أن شركة AT كانـــت تفقد عملاءها بمعدل مقلق. وهذه ظاهرةٌ لها عدة أسباب وأيضًا لها حلولٌ ممكنة. فقد علمت إدارة الشركة أن العديد من الإلغاءات كانت بسبب مكالمات بين العملاء وقســـم خدمة العملاء. ولدراســـة المشكلة بصورة أعمق، تم تشـــكيل لجنة تضم أعضاء من مكتب علاقات العملاء وقسم تقنية المعلومات، وكانت مهمتهم هي النظر في كيفية تقليص مشكلة تحول العملاء بناءً على تحليل أنماط اتصال العملاء (Asamoah وSharda وZadeh وKalgotra، 2016).

عقبات البيانات الضخمة:

عندمـــا يواجه العملاء مشـــكلة في مســـائل معينة مثـــل الفواتير الخاصة بهـــم والخطة وجودة المكالمـــة، فإنهم يتواصلون مع الشركة بطرق متعددة، منهـــا: مركز الاتصال (call center)، وموقع الويب الخاص بالشركة (من خلال روابط اتصل بنا "contact us")، ومركز خدمة الحضور الشخصي (walk-ins). ويمكن للعملاء إلغاء أي حساب من خلال إحدى التفاعلات المذكورة. أرادت الشركة النظر فيما إذا كان تحليل تفاعلات العملاء هذه يمكن أن يُنتج أي أفكار حول أسئلة العملاء أو قناة (قنوات) الاتصال التي استخدموها قبل إلغاء حسابهم. وظهرت البيانات التي تم إنشاؤها بسبب هـــذه التفاعـــلات إما في صورة نصية أو صوتية. ولذا؛ ســـينبغي على شركـــة AT أن تقوم بتجميع كل البيانات في موقع واحد. فقامت الشركة بفحص اســـتخدام المنصات التقليدية لإدارة البيانات، ولكنها اكتشـــفت سريعًا أنها غير كافية لمعالجة تحليل البيانات المتقدم في الســـيناريو الذي يحتوي على تنسيقات متعددة للبيانات من مصادر متعددة (Thusoo وShao وAnthony، 2010).

كان هناك تحديان رئيســـان في تحليل هذه البيانات وهـــما: المصادر المتعددة للبيانات والتي تؤدي إلى مجموعة متنوعة من البيانات وكذلك الحجم الكبير للبيانات.

١- البيانـــات من مصادر متعددة: يســـتطيع العملاء التواصل مـــع الشركة عن طريق الوصول إلى حســـاباتهم على موقع الويب الخاص بالشركة، مما يسمح لشركة AT بإنشاء سجل معلومات على الويب حول نشاط العميل. وقد تمكّنت الشركة باستخدام مسار هذا السجل من تحديد

ما إذا كان العميل قد قام عبر الإنترنت بمراجعة خطته الحالية أو استلامها أو تقديم شكوى بشأنها أو التحقق من الفاتورة وكذلك تحديد متى قام بذلك. وفي مركز خدمة العملاء، يمكن للعملاء أيضًا تسجيل شكوى بشأن الخدمة أو تقديم طلب بتغيير الباقة أو إلغاء الخدمة. سُجِلَتْ هذه الأنشطة في نظام معاملات الشركة وبعد ذلك في مستودع بيانات المنشأة. وأخيرًا، يمكن للعميل الاتصال بمركز خدمة العملاء هاتفيًا وإجراء معاملاته التجارية وكأنه موجود شخصيًا في مركز خدمة العملاء. وكانت سجلات المكالمات متاحة في نظام واحد مع سجل بالأسباب التي يتصل العميل من أجلها. ومن أجل إجراء تحليل له معنى، كان لا بد من تحويل مجموعات البيانات الفردية إلى تنسيقات مهيكلة متشابهة.

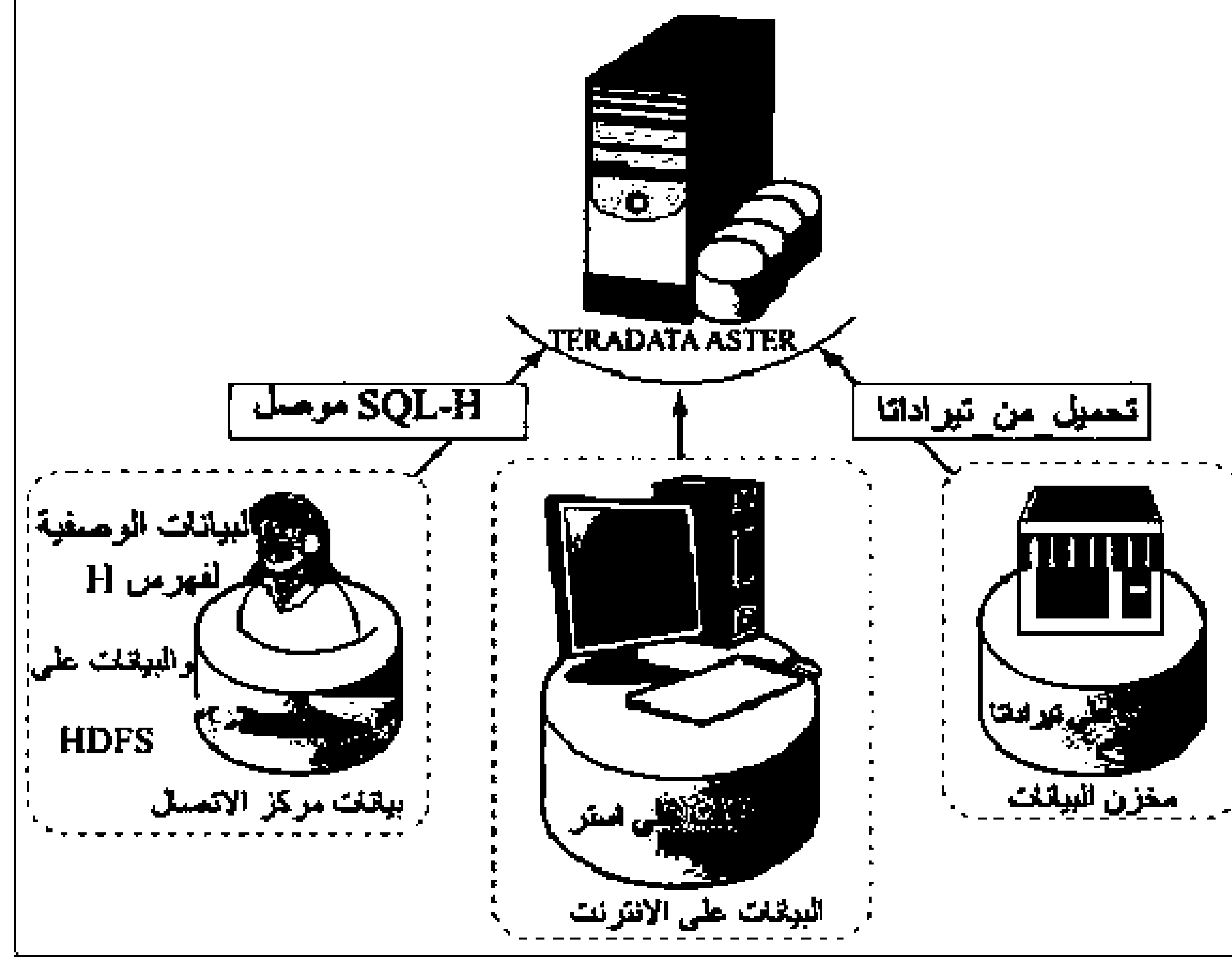
٢- حجم البيانات: كان التحدي الثاني هو الكمية الهائلة من البيانات من المصادر الثلاثة التي استلزم الأمر استخراجها وتنظيفها وإعادة هيكلتها وتحليلها. وبالرغم من أن غالبية مشروعات تحليلات البيانات السابقة اعتمدت على مجموعة عينات صغيرة من البيانات، إلا أن شركة AT قررت الاستفادة من التنوع المتعدد ومصادر البيانات جنبًا إلى جنب مع الحجم الكبير للبيانات المسجلة لتوليد العديد من الأفكار كلما أمكن.

إن النهج التحليلي الذي يمكنه استخدام جميع قنوات ومصادر البيانات، رغم ضخامتها، سيكون قادرًا على توليد أفكار غنية وعميقة من البيانات للمساعدة في تقليل الإلغاءات.

الحل:

تم الاعتماد على بنية موحدة للبيانات الضخمة من شركة تيراداتا أستر (Teradata Aster) لإدارة وتحليل البيانات الضخمة ذات الهيكلة المتعددة. وسنعرض في القسم (٧-٦) بشيء من التفصيل لشركة Teradata Aster. كما يعرض الشكل (٧-١) رسمًا تخطيطيًا للبيانات التي تم تركيبها. وقد تم إنشاء ثلاثة جداول بناءً على كل مصدر للبيانات، حيث يحتوي كل جدول على المتغيرات التالية: رقم العميل، وقناة الاتصال، وختم التاريخ / الوقت، والإجراء المُتَخَذ. وقبل الإلغاء النهائي للخدمة، يمكن لمتغير الإجراء المتخذ أن يكون واحدًا أو أكثر من هذه الخيارات الإحدى عشرة (التي تم تبسيطها لهذه الحالة): تقديم الاعتراض على الفاتورة، أو طلب تحديث الباقة، أو طلب تخفيض الباقة، أو القيام بتحديث الملف الشخصي، أو عرض ملخص الحساب، أو الوصول إلى دعم العملاء، أو عرض الفاتورة، أو مراجعة العقد، أو الوصول إلى وظيفة تحديد موقع المتجر على الويب، أو الوصول إلى قسم الأسئلة المتكررة على الويب، أو استعراض الأجهزة. وقد ركز هدف التحليل على إيجاد المسار الأكثر شيوعًا المؤدي إلى إلغاء الخدمة نهائيًا. وقد تم

تقسيم البيانات إلى مجموعة سلسلة من الأحداث تشتمل على عميل معين في فترة زمنية محددة (٥ أيام على جميع قنوات الاتصال) كجلسة واحدة. وأخيرًا، تم استخدام وظيفة nPath لتسلسل الوقت الخاصة بشركة Aster (والتي تم تفعيلها في إطار SQL-MapReduce) لتحليل الاتجاهات الشائعة التي أدت إلى إلغاء الخدمة.



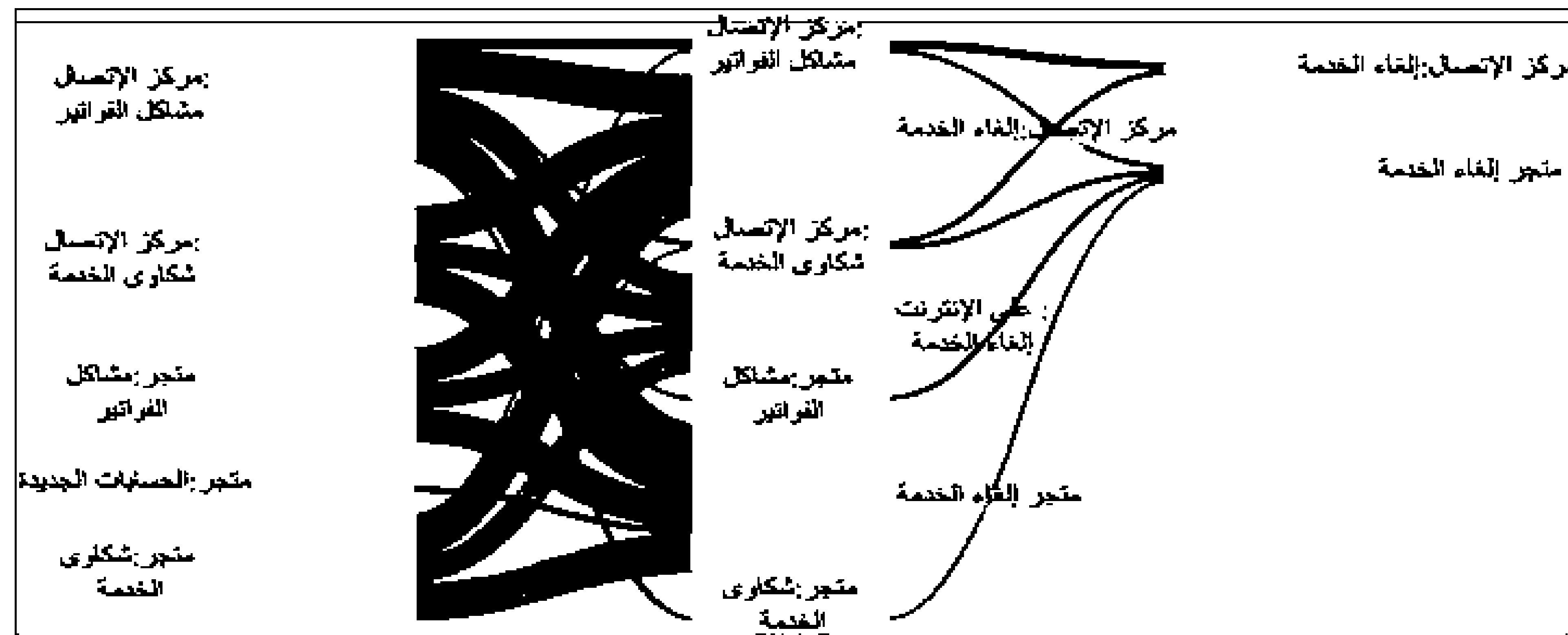
Sources: Teradata Corp.

شكل ٧-١: مصادر بيانات متعددة مدمجة في Teradata Aster

النتائج:

أشـــارت النتائج الأولية إلى العديد من المســـارات التي قد تـــؤدي إلى طلب إلغاء الخدمة، مما مكّن الشركة من تحديد آلاف الطرق التي قد يلجأ إليها العميل لإلغاء الخدمة. وتم إجراء تحليل متابعة لتحديد أكثر الطرق تكرارًا لطلبات الإلغاء. وتم الاصطلاح على هذا الإجراء باســـم المســـار الذهبي "Golden Path". وتم تحديد ٢٠ مسارًا كانوا هم الأكثر تسببًا في الإلغاء في الأجل القصير أو الطويل. ويعرض الشكل (٧-٢) عينة لذلك.

وقد ساعد هذا التحليل الشركة على تحديد العميل قبل إلغاء الخدمة وتقديم حوافز له أو على الأقل تصعيد قرار المشكلة إلى مستوى لا يرتقي إلى إلغاء الخدمة من جانب العميل.



Sources: Teradata Corp.

شكل ٧-٢: التصوير البياني لأعلى عشرين مسارًا للتصوير

ماذا يمكننا أن نتعلم من هذه المقالة القصيرة:

لا تتطلب كل مشاكل الأعمال استخدام منصة تحليلات البيانات الضخمة، بالرغم من أنه يُفضل استخدامها في هذا الموقف. وكان التحدي الرئيسي يدور حول خصائص البيانات محل البحث. وقد مثلت الأنواع الثلاثة المختلفة من مجموعات بيانات تفاعل العميل تحديًا في التحليل. كانت تنسيقات وحقول البيانات المتولدة في كل من هذه الأنظمة ضخمة، وكذلك كان حجم تلك البيانات كبيرًا، مما أدى إلى ضرورة استخدام منصة تستخدم تقنيات تسمح بتحليل حجم هائل من البيانات التي تأتي في تنسيقات متنوعة.

وما يسترعي الانتباه أيضًا هو قيام شركة AT بتوفيق أسئلة البيانات المطروحة على إستراتيجية أعمال المنظمة. وقد أخبرت الأسئلة أيضًا عن نوع التحليل الذي تم إجراؤه. ومن المهم تَفَهُّم أنه بالنسبة لأي تطبيق لبنية البيانات الضخمة، فإن إستراتيجية أعمال المنظمة وتوليد الأسئلة المرتبطة هما مفاتيح تحديد نوع التحليل المراد إجراؤه.

أسئلة مراجعة على المقالة الافتتاحية:

١- ما هي المشكلة التي أحدثها إلغاء خدمة العملاء لاستمرار أعمال شركة AT؟

٢- اذكر مع الشرح العقبات الفنية التي تحدثها طبيعة وخصائص بيانات شركة AT.

٣- ما هو تقسيم البيانات؟ ولماذا تحتم على AT القيام بتقسيم بياناتها؟

٤- قم بإعداد دراسات أخرى تم فيها استخدام نماذج تحول العميل. ما هي أنواع المتغيرات التي تم استخدامها في تلك الدراسات؟ كيف تختلف هذه المقالة الافتتاحية؟

٥- اذكر منصات أخرى شـائعة غـير Teradata Aster، لتحليلات البيانـات الضخمة التي يمكنها التعامل مع التحليل الموضح في الحالة السابقة.

٧-٢ تعريف البيانات الضخمة:

يُشَـكِل اسـتخدام البيانات لفهم الزبائن / العمـلاء وعمليات الأعمال بهـدف الحفاظ على (وتعزيـز) النمـو والربحية تحديًـا كبيرًا ويتزايد هذا التحدي بالنسـبة لمؤسسـات اليوم. وقد أصبحـت معالجة البيانات بالوسـائل التقليدية في حينها غير عملية، وذلك بسـبب توافر المزيد والمزيـد من البيانات في مختلف الأشـكال والأنماط، وتسـمى هذه الظاهـرة في وقتنا الراهن بـ "البيانـات الضخمة"، حيث تحظى بتغطية إعلامية كبيرة بالإضافة إلى اسـتحواذها على اهتمام مستخدمي الأعمال ومتخصصي تقنية المعلومات بشكل متزايد. ونتيجة لذلك؛ فقد أصبح هناك مبالغة وإفراط في استخدام عبارة البيانات الضخمة مما جعلها عبارة طنانة في التسويق.

وبالنسبة لذوي الاهتمامات والخلفيات المختلفة فإن البيانات الضخمة تعني لهم أشياء مختلفة. وقد تم اسـتخدام مصطلح البيانات الضخمة بشكل تقليدي لوصف الكميات الهائلة من البيانات التي تم تحليلها عن طريق منظمات ضخمة مثل جوجل أو مشاريع العلوم البحثية في ناسا. غير أن هذا المصطلح يُعد مصطلحًا نسبيًا لمعظم الأعمال؛ فكلمة "ضخمة" تعتمد على حجم المنظمة. حيث تدور النقطة الأساسـية حول إيجاد قيمة جديدة داخل وخارج مصادر البيانات التقليدية. وتكشـف إزاحة حدود تحليلات البيانات عن إحصاءات وفرص جديدة، وتعتمد كلمة «ضخمة» عـلى المـكان الذي بدأت منه وكيف يسـير تقدمك. وينبغي أن تضع في اعتبـارك الوصف العام للبيانات الضخمة. فالبيانات الضخمة هي التي تتجاوز نطاق بيئات الأجهزة شـائعة الاسـتخدام أو قدرات أدوات البرامج المسـتخدمة لجمعها وإدارتها ومعالجتها في خلال فترة زمنية مقبولة من قِبَلِ المسـتخدمين. وقد أصبح مصطلح البيانات الضخمة مصطلحًا شائعًا لوصف النمو المتسارع، ومدى توفر، واستخدام المعلومات، سواء كان مهيكلًا أم لا. وقد كتب الكثيرون عن اتجاه البيانات الضخمـة وكيـف يمكن أن يكون بمثابة أسـاس للابتكار، والمفاضلة، والنمو. وبسـبب التحديات التقنية في إدارة الحجم الكبير للبيانات الواردة من مصادر متعددة، وأحيانًا بسرعة كبيرة، فقد تم

Sources: Asamoah, D., Sharda, R., Zadeh, A., & Kalgotra, P. (2016). Preparing a big Data analytics professional: A pedagogic experience. In DSI 2016 Conference, Austin, TX. Thusoo, A., Shao, Z., & Anthony, S. (2010). Data warehousing and analytics infrastructure at Facebook. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data (p. 1013). doi: 10.1145/1807167. 1807278.

تطويـر تقنيات جديدة إضافية للتغلب على التحديات التقنيـة. وعادةً ما ترتبط عبارة البيانات الضخمـة بهذه التقنيات. وفي بعض الأحيان يتم توسـيع مصطلـح «البيانات الضخمة» على أنه تحليلات البيانات الضخمة، وذلك بسبب أن الاستخدام الأساسي لتخزين مثل هذه البيانات يؤدي للحصول على إحصاءات من خلال التحليلات. غير أن هذا المصطلح يصبح ذا محتوى مرِن؛ إذ قد يعني أشياء مختلفة لأشخاص مختلفين. وفي هذا الفصل سوف نقوم باستخدام المصطلح الأصلي، لأن هدفنا هو تعريفك بمجموعات البيانات الضخمة وقدراتها في توليد الأفكار.

من أين تأتي البيانات الضخمة؟ أبسـط إجابة عن هذا السـؤال هي أن البيانات الضخمة «في كل مكان». حيث يتم التعامل الآن مع المصادر التي تم تجاهلها بسـبب القيود الفنية على أنها مناجم ذهب. وقد تأتي البيانات الضخمة من سجلات الويب، وتحديد الهوية للموجات اللاسلكية (RFID)، والأنظمة العالمية لتحديد المواقع (GPS)، وشبكات الاستشعار، والشبكات الاجتماعية، والوثائـق النصية المعتمدة عـلى الإنترنت، وفهارس البحث على الإنترنت، والسـجلات التفصيلية للمكالمـات، وعلم الفلك، وعلوم الغلاف الجوي، وعلم الأحياء، وعلم الجينات، والفيزياء النووية، والتجارب البيوكيميائية، والسـجلات الطبية، والبحوث العلمية، والمراقبة العسـكرية، وأرشـيف الصور الفوتوغرافية، وأرشـيف الفيديو، بالإضافة إلى ممارسـات التجـارة الإلكترونية على نطاق واسـع. وليسـت البيانات الضخمة بالشيء الجديد، لكن الجديد هو التغير المسـتمر في تعريف وهيكل البيانات الضخمة. ومنذ ظهور مسـتودعات البيانات في أوائل التسعينيات فإن الشركات تقوم بتخزين وتحليـل كميات ضخمة من البيانات. وعلى الرغم من أن تيرابايتس "terabytes" كان يُسـتخدم كمرادف لمسـتودعات البيانـات الضخمة، إلا أن هذا المصطلح تم اسـتبداله الآن بمصطلح إكسـابايتس "exabytes"، ومـازال معدل النمو في حجم البيانات مسـتمرًا في التصاعد مع سـعي المنظمات إلى تخزين وتحليل مستويات أعلى من تفاصيل المعاملات، فضلًا عن بيانات الويب والميكنة، لاكتساب فهم أفضل لسلوك العملاء وسير الأعمال.

ويعتقـد الكثـير من (الأكاديميـين ومحللي / قادة الصناعـة على حد سـواء) أن إطلاق عبارة «البيانات الضخمة» على البيانات ذات الحجم الهائل هي تسـمية خاطئة. فما تقوله وما تعنيه ليس بالضبط نفس الشيء، بمعنى أن البيانات الضخمة ليسـت فقط «ضخمة». فحجم البيانات الهائل هو خاصية واحدة فقط من ضمن عدة خصائص والتي ترتبط في الغالب بالبيانات الضخمة، وتتضمن هذه الخصائص كلًا من التنوع، والسرعة، والصدق، والتقلب، واقتراح القيمة، وغيرها.

تعريف البيانات الضخمة من خلال مصطلحات "V":

تُعرَّف البيانات الضخمة من خلال المصطلحات الثلاثة للاختصار "V"، وهي: الحجم volume، والتنــوع variety، والسرعــة velocity. وبالإضافة إلى هذه الاختصــارات الثلاثة "V"، فإننا نرى بعــض المقدمــين الرواد لحلول البيانــات الضخمة يضيفون عددًا مــن "V" الأخرى، مثل الصدق veracity (IBM)، والتباين variability (SAS)، بالإضافة إلى اقتراح القيمة value proposition.

الحجم: من الواضح أن الســمة الأكثر شــيوعًا للبيانات الضخمة هي الحجم. وقد ســاهمت عدة عوامــل في الزيادة الهائلة في حجم البيانات، مثل البيانات القائمة على المعاملات والتي تم تخزينها على مر السنين، والبيانات النصية المتدفقة باستمرار من وسائل التواصل الاجتماعي، والكميات المتزايدة من بيانات الاستشعار التي يتم جمعها، بالإضافة إلى بيانات RFID وGPS التي يتم إنشاؤها تلقائيًا، وغيرها. وقد أدى حجم البيانات المفرط في الماضي إلى ظهور مشاكل في التخزين، سواء كانت هذه المشاكل فنية أو مالية. غير أنه حاليًا ومع اســتخدام التقنيات المتطورة والمقترنة بتناقص تكاليف التخزين، فإن هذه المشــكلات لم تعد مهمة؛ وبدلًا من ذلك، فقد ظهرت مشكلات أخرى، ومنها كيفية تحديد الملاءمة فيما بين كميات كبيرة من البيانات وكيفية إنشاء قيمة من البيانات التي تعتبر ذات صلة.

وكما ذكر من قبل، فإن مصطلح ضخمة هو مصطلح نسبي. فهو يتغير بمرور الوقت كما تختلف النظرة إليه باختلاف المنظمات. ومع الزيادة المذهلة في حجم البيانات، فإن تسمية المستوى التالي مــن البيانات الضخمة أيضًا كانت تمثل تحديًا كبيرًا. وقد اســتبدلت أعــلى مجموعة من البيانات والتي كانت تســمى بيتابايتس petabytes (PB) ليحل محلها زيتابايتس (ZB)، والتي هي عبارة عــن تريليون جيجابيت (GB) أو مليار تيرابايــت (TB). وتقدم رؤية فنية (٧-١) لمحة عامة عن حجم وتسمية وحدات مستودعات البيانات الضخمة.

ومــن منظور تاريخــي قريب فإن العالم كان يملك حوالي ٠.٨ ZB مــن البيانات في عام ٢٠٠٩؛ وبحلول عام ٢٠١٠، تخطت هذه المجموعة حاجز الـ ١ ZB؛ وفي نهاية عام ٢٠١١، كان العدد قد وصل إلى ١,٨ ZB. ومن المتوقع أن يصل العدد إلى ٤٤ ZB بحلول عام ٢٠٢٠ (Adshead, 2014). ومع نمو أجهزة الاستشعار وشبكة الإنترنت للأشياء (IOT - والتي سنقدمها في الفصل التالي)، فإن هذه التوقعات قد تكون خاطئة. وعلى الرغم من أن هذه الأرقام مذهلة من حيث الحجم، فإن التحديات والفرص التي تأتي معها هي أيضًا كذلك.

رؤية فنية ٧-١: حجم البيانات يتضخم بشكل أكبر وأكبر

إن قياس حجم البيانات يواجه صعوبة في مواكبة الأسماء الجديدة. فنحن جميعًا نعرف الكيلوبايت (KB، وهــو ١,٠٠٠ بايت)، والميجابايــت (MB، وهي مليون بايت)، والجيجابايت (GB، وهي مليار

بايت)، والتيرابايت (TB، وهي تريليون بايت). وخلاف ذلك، فإن الأسـماء المعطاة لأحجام البيانات جديدة نسبيًّا بالنسبة لمعظمنا. ويوضح الجدول التالي ما يأتي بعد التيرابايت وما بعده.

الاسم	الرمز	القيمة
كيلوبايت	KB	$١٠^{٣}$
ميجابايت	MB	$١٠^{٦}$
جيجابايت	GB	$١٠^{٩}$
تيرابايت	TB	$١٠^{١٢}$
بيتابايت	PB	$١٠^{١٥}$
إكسابايت	EB	$١٠^{١٨}$
زيتابايت	ZB	$١٠^{٢١}$
يوتابايت	YB	$١٠^{٢٤}$
برونتوبايت*	BB	$١٠^{٢٧}$
جيجوبايت*	GeB	$١٠^{٣٠}$

* اسم / رمز لم يعتمد رسميًّا حتى الآن من قِبَل SI (النظام الدولي للوحدات).

ضـع في اعتبارك أنه يتم إنشـاء إكسـابايت من البيانات على الإنترنت كل يـوم، والتي تعادل ٢٥٠ مليـون قـرص فيديو رقمي DVDs من المعلومات ذات القيمة. وعندما يتعلق الأمر بكمية المعلومـات التـي تمتد عبر الويـب كل عام، فإن فكـرة وجود كميات أكبر مـن البيانات - وهي زيتابايت - ليسـت بعيدة جدًا. وفي الحقيقة، فإن خبراء الصناعة يقدرون بالفعل أننا سنشـاهد سـنويًّا عـبر الإنترنت ١,٣ زيتابايت من حركة مرور البيانـات بحلول عام ٢٠١٦ - ومن الممكن أن تقفـز لتصل إلى ٢,٣ زيتابايت بحلول عام ٢٠٢٠. ومن المتوقع أن تصل زيارات الإنترنت إلى ٣٠٠ جيجابايت لكل فرد سـنويًّا بحلول عام ٢٠٢٠. وعند الإشـارة إلى يوتابايت، فإنه غالبًا ما يتساءل بعض علماء البيانات الضخمة عن مقدار البيانات التي لدى NSA أو FBI عن الناس إجمالًا. كما أن وضع الشروط الخاصة بأقراص الفيديو الرقمية DVDs، يوتابايت سوف يتطلب ٢٥٠ تريليون منهـم. أما يوتابايت، والتي هي عبارة عن ١ متبوعًا بـــ ٢٧ صفرًا، فإنها إلى الآن لم يتم اعتمادها من SI الرسمية غير أنه من الواضح أنها معترف بها من قِبل بعض الناس في مجتمع القياس، حيث يمكن اسـتخدام هذا الحجم من الضخامة لوصف مقدار بيانات الاستشـعار التي سـوف نحصل عليها من الإنترنت في العقد القادم، إن لم يكن قبل ذلك.

أما بالنسبة لـ جيجوبايت فهي تمثل ١٠ مرفوعة إلى الأس ٣٠. وفيما يتعلق بالمصدر الذي تأتي منه البيانات الضخمة، فإنه يجب أخذ ما يلي في الاعتبار:

- يقوم مصادم الهدرونات الكبير CERN بتوليد ١ بيتابايت في الثانية الواحدة.
- تنتج أجهزة الاستشعار من محرك طائرة بوينغ ٢٠ تيرابايت من البيانات كل ساعة.
- يتم استيعاب ٦٠٠ تيرابايت من البيانات الجديدة كل يوم في قواعد بيانات Facebook.
- يتم تنزيل ٣٠٠ تسجيل من تسجيلات الفيديو في الدقيقة، مع ترجمة ١ تيرابايت كل دقيقة على اليوتيوب.
- سـوف يقوم التلسـكوب المقترح لمقياس الكيلومتر المربع (وهو أكبر تلسكوب في العالم) بتوليد إكسابايت من البيانات يوميًا.

التنـوع: تتشـكل البيانات في الوقت الحاضر بجميع أنواع التنسـيقات - بدءًا مـن قواعد البيانات التقليدية إلى مخازن البيانات الهرمية والتي أُنشـئَت من قِبل المستخدمين النهائيين بالإضافة إلى أنظمة المعالجة التحليلية الفورية، والوثائق النصية، والبريد الإلكتروني، وXML، والبيانات التي جُمِعَت بواسطة جهاز استشـعار، وبيانات الاستشـعار، فضلًا عن بيانـات الفيديو، والصـوت، وبيانات شريط مؤشرات الأسـهم. وتشـير بعض التقديرات إلى أن ٨٠ إلى ٨٥٪ من بيانات جميع المنظمات تأخذ أحد الأشـكال التنظيمية المهيكلة أو شبه المهيكلة (وهو تنسيق غير مناسب لمخططات قواعد البيانات التقليدية)، غير أنه لا يمكن إنكار قيمته، وبالتالي، فإنه ينبغي تضمينه في التحليلات من أجل دعم القرار.

السرعـة: وفقًا لمـا ذكره Gartner، فإن السرعة يُقصد بها كلٌ من السرعة في إنتاج البيانات ومدى سرعة معالجة البيانات (بمعنى أن يتم التقاطها وتخزينها وتحليلها) من أجل تلبية الحاجة أو الطلب. وتقـوم بطاقـات التعريف اللاسـلكية RFID وأجهزة الاستشـعار الآلية بالإضافـة إلى أجهزة GPS والعدادات الذكية بالعمل على التوجه بالاحتياج المتزايد للتعامل مع سيول البيانات في الوقت المناسب تقريبًـا. وربما تكون السرعة هي الخاصية الأكثر إغفالًا من خواص البيانات الضخمة؛ حيث إن إعطاء رد الفعـل بسرعة كافيـة للتعامل مع السرعة يمثل تحديًا لمعظم المنظمات. وبالنسـبة للبيئات ذات الحساسية تجاه الوقت، فإن ساعة التكلفة البديلة للبيانات تبدأ بتحديد اللحظة التي يتم فيها إنشاء البيانـات. ومـع مرور الوقت، تتراجع القيمة المقترحة للبيانـات إلى أن تصبح في نهاية المطاف عديمة القيمـة. ويؤدي الوصول إلى البيانات والاسـتجابة للظروف بشـكل أسرع دائمًا إلى نتائج أكثر فائدة، سواء أكان الموضوع يتعلق بصحة المريض، أو سلامة نظام المرور، أو صحة السندات الاستثمارية.

وفي عاصفة البيانات الضخمة التي نشـهدها حاليًا، فإن جميع الأشـخاص تقريبًا يركزون على تحليلات حالة الراحة at-rest analytics، وذلك باستخدام أنظمة برامج وأجهزة مُحَسَّنة لتجميع

كميـــات كبـــيرة من مصادر البيانات المختلفة. وعلى الرغم من أن هذا الأمر مهم للغاية فضلًا عن قيمته العالية، إلا أن هناك فئة أخرى من التحليلات، مدفوعة بسرعة البيانات الضخمة، وتســـمى «تحليـــلات تدفق البيانات» أو «التحليلات في حالة الحركة»، والتي تتطور بسرعة. فإذا تم القيام بتحليلات تدفق البيانات بشكل صحيح ففي هذه الحالة يمكنها أن تكون ذات قيمة كبيرة، بل إنها قد تكون أكثر قيمة في بعض بيئات العمل من تحليلات حالة الراحة at-rest analytics. وسوف نتناول هذا الموضوع بشيء من التفصيل لاحقًا في هذا الفصل.

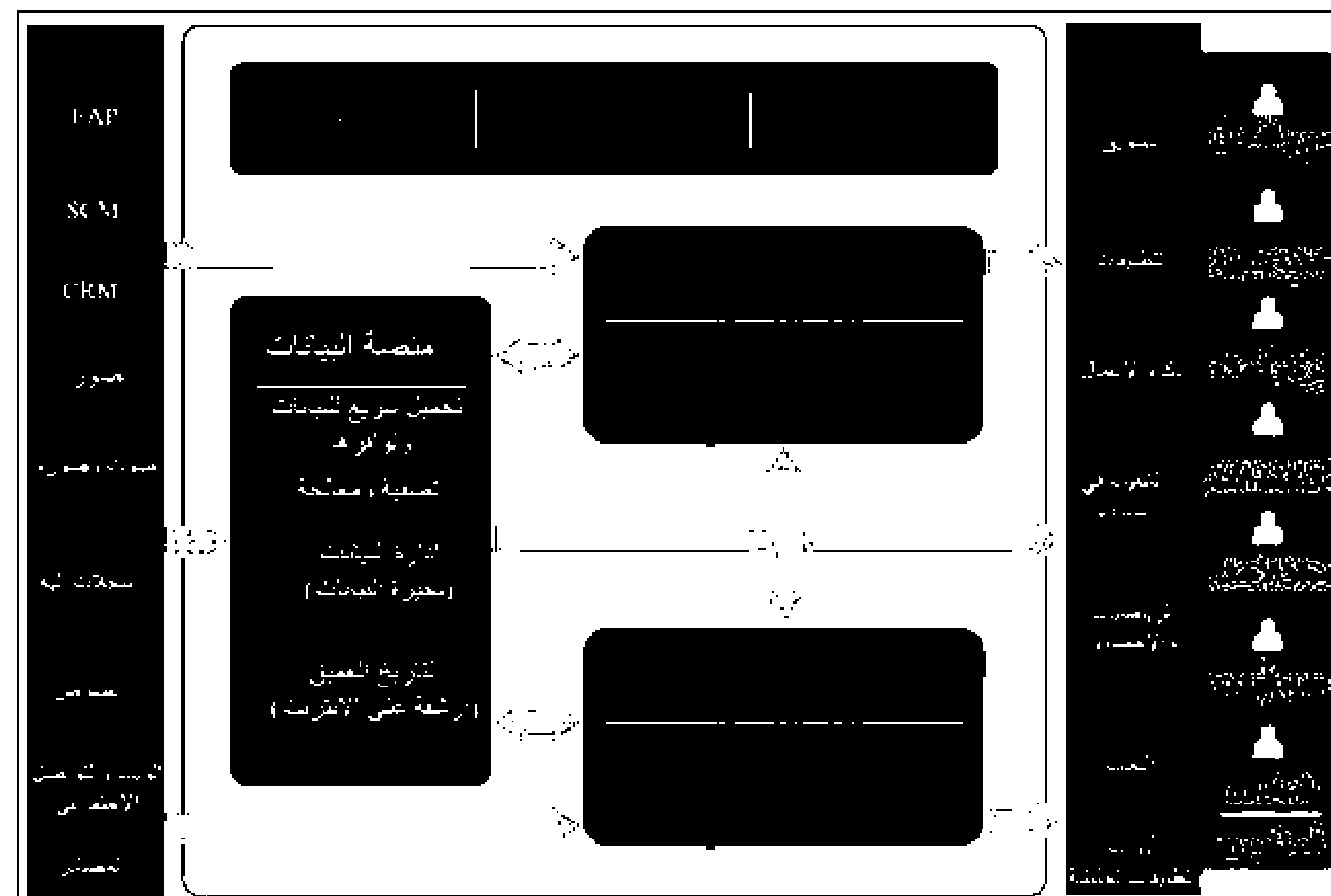
الصدق: الصدق هو مصطلح تم صياغته بواسطة IBM والذي يُستَخدم على أنه الـ "V" الرابعة لوصف البيانات الضخمة. ويشـــير هذا المصطلـــح إلى التطابق مع الحقائق: الدقة، أو الجودة، أو الصـــدق، أو الثقة في البيانات. وغالبًا ما يتم اســـتخدام الأدوات والأســـاليب من أجل التعامل مع صحة البيانات الضخمة وذلك عن طريق تحويل إلى بيانات ذات جودة عالية وجديرة بالثقة.

القابلية للتغيير: فضلًا عن تزايد السرعات والتنوع في البيانات، يمكن أن تكون تدفقات البيانات غـــير متســـقة بدرجة كبيرة مع القمـــم الدورية. هل هناك شيء كبير يتجه نحو وســـائل التواصل الاجتماعي؟ ربما يكون هناك IPO (طرح عام مبدئي) رفيع المستوى يلوح في الأفق. وربما تصبح الســـباحة مع الخنازير في جزر البهاما فجأة نشاطًا واجب الأداء أثناء العطلة. كما يمكن أن تكون الأحمال القصوى للبيانات اليومية والموسمية والناجمة عن الحدث متغيرة إلى حد كبير مما يُشَكل تحديًا للإدارة - خاصة مع وسائل التواصل الاجتماعي المعنية.

عـــرض القيمة: تتمثل الإثارة حول البيانات الضخمة في عروض القيمة الخاصة بها. أما الفكرة المسبقة حول البيانات «الضخمة» فتتمثل في كونها تحتوي (أو لديها قدرة أكبر على احتواء) المزيد من الأنماط والأشكال المثيرة للاهتمام أكثر من البيانات «الصغيرة». وبالتالي، فإن المنظمات يمكنها أن تكتســـب قيمة تجارية أكبر من خلال تحليل البيانـــات الضخمة والغنية بالميزات، والتي قد لا يمكنها أن تكتســـبها بطريقة أخرى. وعلى الرغم من أن المســـتخدمين يمكنهم اكتشاف الأنماط في مجموعات البيانات الصغيرة من خلال اســـتخدام أســـاليب إحصائية بسيطة وطرق تعلُّم الآلة أو أدوات طلب البحث وإعداد التقارير المخصصة، إلا أن البيانات الضخمة تعني تحليلات «ضخمة». وتعني التحليلات الضخمة رؤية أكبر وأشياء وقرارات أفضل، والتي تحتاج إليها كل منظمة.

ومـــن المرجح أن تتم إضافـــة المزيد من الخصائص (ربما تكون المزيد مـــن "V") إلى تلك القائمة، وذلك نظرًا لكون التعريف الدقيق للبيانات الضخمة (أو بنودها اللاحقة) لازال مطروحًا للمناقشـــة المستمرة في الدوائر الأكاديمية والصناعية. وبغض النظر عما يحدث، فإن أهمية وعرض قيمة البيانات الضخمة وُجِدَت لتبقى. ويعرض الشـــكل (٧-٣) بنية مفاهيمية حيث يتم تحويل البيانات الضخمة

(الموجودة في الجانب الأيسر من الشـكل) إلى رؤية عملية من خلال اسـتخدام مزيج من التحليلات المتقدمة وتقديمها لمجموعة متنوعة من المستخدمين / الأدوار المختلفة لاتخاذ قرارات أسرع / أفضل.



SOURCE: ASTERDATA–A TERADATA COMPANY.

شكل ٧-٣: بنية مفاهيمية عالية المستوى لحلول البيانات الضخمة

وهناك مصطلح آخر تتم إضافته إلى الكلمات الطنانة للبيانات الضخمة وهو مصطلح البيانات البديلـة alternative data. وتعـرض الحالة العملية (٧-١) أمثلة لأنواع متعددة من البيانات في عدد من السيناريوهات المختلفة.

حالة عملية ٧-١

بيانات بديلة لتحليلات أو توقعات السوق

يُعد الحصول على توقعات جيدة وفهم الوضع بشـكل جيد أمرًا ضروريًّا لأي سيناريو، إلا أنه مهم بشـكل خاص لممثلي صناعة الاستثمار. كما أن إمكانية حصول المستثمر على مـؤشر مبكر لكيفية عمل مبيعات معينة لمتاجر التجزئة من الممكن أن تمنحه قائمة حول شراء أو بيع أسـهم هذا البائع حتى قبل أن تصدر تقارير الأرباح. وتُعد مشـكلة التنبؤ

Sources: Higginbotham, S. (2012). As data gets bigger, what comes after a yottabyte? gigaom.com/2012/10/30/as-data-gets-bigger- what-comes-after-a-yottabyte (accessed August 2016). Cisco. (2016). The zettabyte era: Trends and analysis. cisco.com/c/en/us/solutions/collateral/service-provider/visualnetworking- index-vni/vni-hyperconnectivity-wp.pdf (accessed August 2016).

بالنشاط الاقتصادي أو المناخ المحلي الذي يعتمد على مجموعة متنوعة من البيانات بخلاف بيانات التجزئة المعتادة هي ظاهرة حديثة جدًا كما أنها أدت إلى ظهور عبارة أخرى طنانة وهي «البيانات البديلة». ويُعد المزيج الرئيس في هذه الفئة من البيانات البديلة هو صور الأقمار الصناعية، غير أنه يتضمن أيضًا بيانات أخرى مثل وسائل التواصل الاجتماعي، أو الإيداعات الحكومية، أو إعلانات الوظائف، أو أنماط حركة المرور، أو التغييرات في مواقف السيارات أو المساحات المفتوحة التي اكتُشِفَت من خلال صور الأقمار الصناعية، بالإضافة إلى أنماط استخدام الهاتف المحمول في أي موقع محدد خلال أي وقت محدد، وأنماط البحث على محركات البحث، وغير ذلك. وقد قامت شركة فيسبوك وغيرها من الشركات بالاستثمار في الأقمار الصناعية محاولةً تصوير العالم كله بشكل يومي بحيث تتمكن من تعقب التغييرات اليومية في أي مكان ومن ثم استخدام المعلومات من أجل التنبؤ. وقد ورد العديد من الأمثلة المثيرة للاهتمام لتنبؤات متقدمة وموثوقة بدرجة عالية. وفي الواقع، فإن هذا النشاط تقوده شركات ناشئة. وفيما يلي بعض الأمثلة:

- استخدام Facebook محرك التعرف على الصور من أجل تحليل أكثر من ١٤,٦ مليار صورة بغرض تحليل كل ركن من أركان العالم لتحديد المناطق ذات الاتصال المنخفض.

- راقبت شركة RS Metrics مواقف السيارات عبر الولايات المتحدة من خلال صناديق الاستثمار المختلفة، حيث توقعت شركة RS Metrics تحقيق أرباح قوية في الربع الثاني من عام ٢٠١٥ لصالح JC Penney، وقد اعتمدت الشركة في توقعاتها على تحليل مواقف السيارات. وقد استفاد عملاؤها (في الغالب عملاء صناديق الاستثمار) من هذه الرؤية المتقدمة. وقد وردت قصة أخرى مشابهة لـ Wal-Mart والتي استخدمت أعداد السيارات الموجودة في مواقف السيارات الخاصة بها من أجل توقع المبيعات.

- قيام شركة Orbital Insights بتوفير مؤشرات الاقتصاد الكلي لقطاعات الصناعة المختلفة من خلال استخدام بيانات صور الأقمار الصناعية. فعلى سبيل المثال، تَدَّعِي الشركة أنها أنتجت تقديرات يومية أفضل مما هو متاح الآن من وكالة الطاقة الدولية (IEA) من أجل تخزين النفط في جميع أنحاء العالم، وقد استطاعت فعل ذلك من خلال تحليل صهاريج تخزين النفط حول العالم.

- تتبع شركة Spaceknow للتغيرات في محيط المصنع لأكثر من ٦٠٠٠ موقع مصنع صيني. وقد تمكنت الشركة باستخدام هذه البيانات، من تقديم فكرة عن النشاط الاقتصادي الصناعي للصين، وقد كانت هذه الفكرة أفضل بكثير مما قدمته الحكومة الصينية.

- اســتخدام معامل Descartes لبيانات الأقمار الصناعية من أجل التنبؤ بحصاد الذُّرة في الولايــات المتحدة بدقة أكبر من وزارة الزراعــة الأمريكية. فالتنبؤات الأفضل قد يكون لها آثار مالية ضخمة على تداول العقود الآجلة. ومن أقدم الأمثلة على ذلك مثال لإحدى الشركات تُسمى Lanworth والتي تنبأت أيضًا بتقديرات محاصيل الذرة. وقد قامت شركة Thomson Reuters بشراء شركة Lanworth ثم دمجتها في خدمة Eikon الخاصة بهم.

- قــدرة شركة DigitalGlobe على تحليــل حجم الغابة بمزيد من الدقة نظرًا لقدرة برنامجها على عد كل شــجرة في الغابة، مما يؤدي إلى تقدير أكثر دقة نظرًا لأنه لا توجد حاجة لاستخدام عينة تمثيلية.

- قيام شركة Kensho بالاســتفادة من تحليل البيانــات من مصادر متعددة (ذُكِرت سابقًا) من أجل بناء محرك تجاري، وذلك بدعم من شركة Goldman Sachs.

وقــد أعطت هذه الأمثلة مجرد عينة من الطرق التي يمكن بها جمع البيانات من أجل إنشــاء أفكار جديــدة. وبالطبع، فإنه في بعض الحالات يكــون هناك بعض المخاوف التي تتعلق بالخصوصية. فعلى ســبيل المثــال، ذكرت صحيفة Wall Street في عام ٢٠١٥ قصة لشركة Yodlee، وهي شركة توفر أدوات التمويل الشــخصي لعدد من البنوك الكبيرة ومن ثــم فإن الشركة يمكنها الوصول إلى ملايين من معاملات بطاقات الائتمان الخاصة بالعملاء، وتقوم ببيع هذه البيانات إلى شركات تحليلات أخرى والتي يمكنها استخدام هذه المعلومات من أجل تطوير تنبؤات مبكرة حول كيفية اتجاه المبيعات نحو بائع تجزئة معين.

وهذه المعلومات مطلوبة بشــكل كبير من قِبل متعاملي ســوق الأوراق المالية. وقد أدت هذه القصة إلى حدوث ضجة كبيرة بســبب استخدام المعلومات الخاصة بالعملاء بطرق غير مصرح بها. أيضًا، فإن هناك قلقًا في بعض الدوائر حول مشروعية وضع مثل هذه التنبؤات المتقدمة حول ســلعة معينة أو شركة معينة. وعلى الرغم من أن جميع هذه المخاوف سيتم حلها في نهاية المطاف من قِبل صانعي السياسات، إلا أنه من الواضح أن الطرق الجديدة والمثيرة للاهتمام والتي تُستَخدم للجمع بين بيانات الأقمار الصناعية والعديــد من مصادر البيانــات الأخرى ينتج عنها ظهور مجموعــة جديدة من شركات التحليلات، حيــث تعمل جميع هذه المنظمات مع البيانات التي تتوافق مع مواصفات الثــلاث (V's) وهي تنوع العناصر وحجمها وسرعاتهــا. وبالإضافة إلى ذلك، فإن بعض هــذه الشركات تعمــل أيضًا مع فئة أخرى من البيانات وهي أجهزة الاستشــعار، والتي سنقوم بمناقشتها في الفصل التالي عندما نستعرض الاتجاهات الناشئة في التحليلات. غير أن هذه الشركات تقع ضمن مجموعة من التطبيقات المبتكرة والناشئة.

أسئلة للمناقشة:

١- ما هو العامل المشترك في الأمثلة التي تناولناها بالمناقشة في هذه الحالة العملية؟

٢- هــل يمكنك التفكير في تدفقات البيانات الأخرى التي قد تســاعد في إعطاء دلالة مبكرة للمبيعات في متاجر التجزئة؟

٣- هل يمكنك التفكير في تطبيقات أخرى تحاكي الطرق التي تم عرضها في هذه الحالة العملية؟

Sources: Dillow, C. (2016). What happens when you combine artificial intelligence and satellite imagery. fortune.com/201630/03//facebook-ai-satellite-imagery/ (accessed July 2016). Ekster, G. (2015). Driving investment performance with alternative data. integrity-research.com/wp-content/uploads/201511//Driving-Investment-Performance-With-Alternative-Data.pdf (accessed July 2016). Hope, B. (2015). Provider of personal finance tools tracks bank cards, sells data to investors. wsj.com/articles/providerof-personal-finance-tools-tracks-bank-cards-sells-data-to-investors-1438914620 (accessed July 2016). Orbital Insight. World Oil Storage Index. orbitalinsight.com/solutions/world-oil-storage-index/(accessed July 2016). Shaw, C. (2016). Satellite companies moving markets. quandl.com/blog/alternative-data-satellite-companies (accessed July 2016). Steiner, C. (2009). Sky high tips for crop traders. http://www.forbes.com/forbes/20090907//technology-software-satellites-sky-high-tipsfor-crop-traders.html (accessed July 2016). Turner, M. (2015). This is the future of investing, and you probably can't afford it. businessinsider.com/hedge-funds-are-analysing-data-to-get-an-edge-20158- (accessed July 2016).

أسئلة للمراجعة على القسم ٧-٢:

١- لماذا تعتبر البيانات الضخمة مهمة؟ وما الذي تغير ليتم وضعها في مركز عالم التحليلات؟

٢- عرِّف البيانات الضخمة. ولماذا يصعب تعريفها؟

٣- ما هو أصل "V's" التي تُســتخدم لتعريف البيانات الضخمة، وأي منها يعتبر الأكثر أهمية في رأيك؟ ولماذا؟

٤- ما رأيك في مستقبل البيانات الضخمة؟ وهل من الممكن أن تترك شهرتها لشيء آخر؟ وإذا كان الأمر كذلك، فماذا سيكون هذا الشيء؟

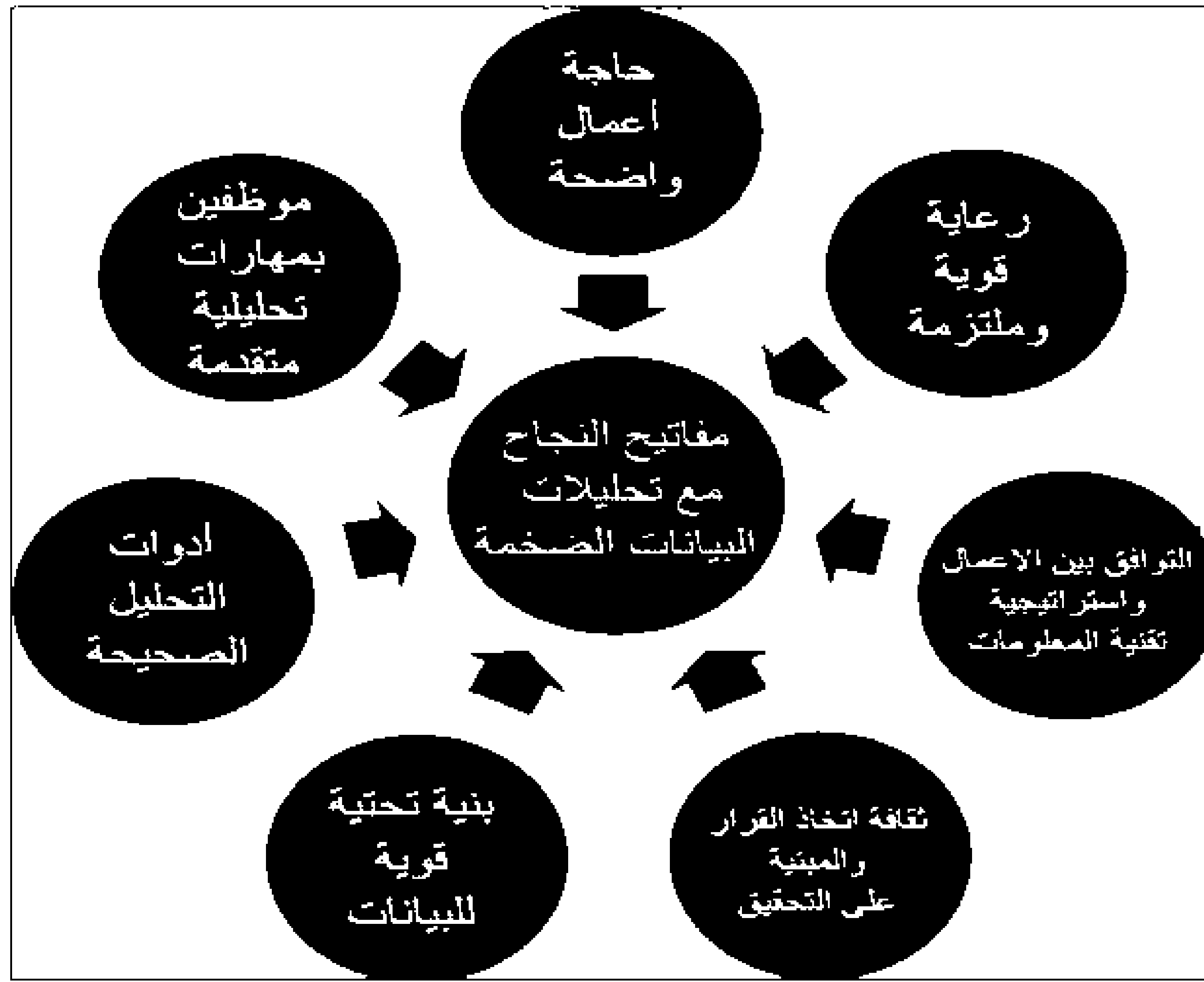
٧-٣ أساسيات تحليلات البيانات الضخمة:

لا قيمــة للبيانــات الضخمة بحــد ذاتها - بغض النظر عــن حجم هذه البيانــات، أو نوعها، أو سرعتها - مالم يقم مستخدمو الأعمال بعمل شيء ما يعطي قيمة لمنظماتهم. وهنا تظهر في الصورة التحليلات "الضخمة". وعلى الرغم من قيام المنظمات دائمًا بتشــغيل التقارير ولوحات المعلومات ضد مســتودعات البيانات، إلا أن معظمها لم يفتح هذه المســتودعات لإجراء استكشــاف متعمق حســب الطلب، وهذا يرجع بشــكل جزئي إلى التعقيد الشديد لأدوات التحليل بالنسبة للمستخدم العــادي، كــما يرجع أيضًا إلى أن المســتودعات في الغالــب لا تحتوي على جميــع البيانات اللازمة للمســتخدم القوي، غير أن كل ذلك على وشــك التغيير بطريقة مثيرة (وبالفعل تم تغيير البعض)، وذلك بفضل النموذج الجديد لتحليلات البيانات الضخمة.

وعلى الرغم من عرض قيمة البيانات الضخمة، غير أنها تسببت أيضًا في تحديات كبيرة للمنظمات، حيــث إن الوســائل التقليدية لالتقاط وتخزيــن وتحليل البيانات غير قادرة عــلى التعامل بفاعلية وكفاءة مع البيانات الضخمة. ولذلك؛ فإن السلالات الجديدة للتقنيات ينبغي تطويرها (أو شراؤها / استخدامها / الاستعانة بمصادر خارجية) من أجل مواجهة تحديات البيانات الضخمة. كما ينبغي على المنظمات قبل اتخاذ أي خطوات في هذا الاستثمار أن تبين أسباب استخدام مثل تلك الوسائل. وفيما يلي نذكر بعض الأمثلة التي قد تساعد في تسليط الضوء على هذه الحالة. واعلم أنك ستكون بحاجة للتفكير الجاد في الإبحار في رحلة البيانات الضخمة في حال تحقُّق إحدى الحالات التالية:

- إذا كنــت لا تســتطيع معالجة مقدار البيانات التي تريدها بســبب المحــددات المفروضة من نظامك الحالي أو البيئة الحالية.
- إذا كنــت تريــد تضمين مصادر بيانــات جديدة / معاصرة (مثل: وســائل التواصل الاجتماعي، RFID، الحســيّة، الويــب، GPS، البيانــات النصية) في منصة البيانــات الخاصة بك، غير أنك لا تستطيع ذلك نظرًا لعدم توافقها مع الصفوف والأعمدة المحددة في مخطط مستودعات البيانات دون التضحية بثبات أو ثراء البيانات الجديدة.
- أن تكون في حاجة إلى (أو ترغب في) دمج البيانات بأسرع ما يمكن لتتوافق مع تحليلاتك.
- أن تكون لديك رغبة في العمل بمخطط حســب الطلب (بدلًا من المخطط المحدد ســلفًا والذي يُستخدم في أنظمة إدارة قواعد البيانات ذات الصلة [RDBMSs]) لأن طبيعة البيانات الجديدة قد لا تكون معروفة، أو قد لا يكون هناك متسع من الوقت لتحديد ذلك وتطوير المخطط له.
- وصول البيانات إلى منظمتك بصورة سريعة، بحيث لا تستطيع منصة التحليلات التقليدية التعامل معها.

وكما هو الحال مع أي استثمار كبير في مجال تقنية المعلومات، فإن النجاح في تحليلات البيانات الضخمــة يعتمد على عــدد كبير من العوامل. ويعرض الشــكل (٧-٤) تصويرًا بيانيًا لأكثر عوامل النجاح أهمية (Watson, 2012).



شكل ٧-٤: عوامل النجاح لتحليلات البيانات الضخمة

وفيــما يــلي أهــم عوامــل نجــاح تحليــلات البيانــات الضخمــة (Watson, Sharda, & Schrader, 2012):

١- وضوح احتياجات الأعمال (بحيث تتوافق مع الرؤية والإستراتيجية): فالاستثمارات في الأعمال ينبغــي أن تتم لصالــح العمل، وليس لمجرد التقــدم بمجال التقنية. ولذلــك؛ فإنه ينبغي أن تكون احتياجات الأعمال هي المحرك الرئيس لتحليلات البيانات الضخمة على كل المستويات: الإستراتيجي، والتكتيكي، بالإضافة إلى التشغيلي.

٢- رعاية قوية، ملتزمة (بطل تنفيذي): فمن المعروف جيدًا أنه من الصعب (إن لم يكن مستحيلًا) تحقيــق النجــاح، إذا لم يكن لديك رعاية ودعم تنفيذي قــوي وملتزم. هذه الرعاية يمكن أن تكون على مســتوى الإدارات، وذلك في حالة إذا كان المدى عبارة عن تطبيق تحليلي فردي أو عــدد قليل من التطبيقات التحليلية. ومــع ذلك، فإنه في حالة ما إذا كان الهدف هو التحول التنظيمي على نطاق المؤسسة، وهو ما يحدث في كثير من الأحيان لمبادرات البيانات الضخمة، فينبغي أن تكون الرعاية على أعلى المستويات كما ينبغي أن يكون التنظيم واسع النطاق.

٣- التوافـق بين الأعمال وإسـتراتيجية تقنية المعلومـات: فمن الضروري التأكـد من أن العمل التحليلي يدعم دائمًا إسـتراتيجية العمل وليس العكس. وينبغـي أن تلعب التحليلات الدور التمكيني في تنفيذ إستراتيجية الأعمال بنجاح.

٤- ثقافة صنع القرار المبني على الحقيقة: في ثقافة صنع القرار المبني على الحقيقة، يكون الاعتماد على الأرقام بدلًا من الحدس، أو الشـعور الغريزي، أو افتراض عملية صنع القرار. وهناك أيضًا ثقافة التجريب لمعرفة ما يصلح ومالا يصلح. ولإنشاء هذه الثقافة تحتاج الإدارة العليا إلى:

- إدارك أن بعض الأشخاص لا يمكنهم أو لن يقوموا بضبط ذلك.
- أن تكون داعمًا صوتيًا.
- التأكيد على ضرورة وقف الأساليب القديمة.
- أن تطلب معرفة أي التحليلات دخلت في القرار.
- أن تربط الحوافز والمكافآت بالسلوكيات المرغوبة.

٥- بنيـة تحتية قوية للبيانات: قدمت مسـتودعات البيانات البنية الأساسـية للبيانات من أجل التحليـلات. وفي عـصر البيانات الضخمة فإن هذه البنية الأساسـية يتـم تغييرها وتعزيزها باستخدام تقنيات جديدة، حيث يتطلب النجاح أن يتم الجمع بين القديم والجديد من أجل بنية أساسية شاملة تعمل بصورة تآزرية.

وتـزداد الحاجة إلى أنظمة تحليلية أكثر كفاءة، كلـما ازداد الحجم والتعقيد. وقد تم تطوير عدد من التقنيات الحسابية والمنصات المبتكرة والجديدة، من أجل مواكبة الاحتياجات الحسابية للبيانات الضخمة. وهذه التقنيات مجتمعةً تسمى الحوسبة عالية الأداء، والتي تتضمن ما يلي:

- تحليـلات داخل الذاكرة: وهي تقوم بحل المشـاكل المُعقدة فور حدوثهـا تقريبًا بأفكار عالية الدقة من خلال السماح بمعالجة وتوزيع الحسابات التحليلية والبيانات الضخمة في الذاكرة عبر مجموعة مخصصة من العقد.
- تحليلات داخل قاعدة البيانات: وتعمل على توفير الوقت للأفكار وتمكين إدارة البيانات بشكل أفضـل من خلال إجراء تكامل للبيانـات والوظائف التحليلية داخل قاعـدة البيانات بحيث لا تضطر إلى نقل البيانات أو تحويلها بشكل متكرر.
- الحوسبة الشبكية: وتعمل على رفع الكفاءة، وخفض التكلفة، وتحسين الأداء عن طريق معالجة الوظائف بشكل مشترك تتم إدارته مركزيًا من قِبل موارد تقنية المعلومات.
- الوسائل: يتم الجمع بين الأجهزة والبرامج في وحدة مادية لا تتسم بالسرعة فقط، بل هي أيضًا قابلة للتطوير على حسب الحاجة.

وتمثـل المتطلبات الحاسـوبية مجـرد جزء صغير من قائمـة التحديات التـي تفرضها البيانات الضخمـة على مشروعات اليوم. وقد عثر مديرو الأعمال التنفيذيون على تحديات ذات تأثير كبير على التنفيذ الناجح لتحليلات البيانات الضخمة. وعند النظر في مشـاريع وأسـاليب بناء البيانات الضخمة، فإن التنبه لهذه التحديات سـوف يجعل رحلة تحليلات الكفاءة أقل إرهاقًا. وفيما يلي قائمة بهذه التحديات:

- حجـم البيانات: القدرة على جمع، وتخزيـن، ومعالجة حجم ضخم من البيانات بسرعة ملائمة حيث تكون أحدث المعلومات متاحة لصانعي القرار في حالة حاجتهم إليها.
- تكامل البيانات: القدرة على الجمع بين البيانات غير متشـابهة الهيكل أو المصدر مع مراعاة أن يتم عمل ذلك بسرعة وبتكلفة معتدلة.
- إمكانـات المعالجة: القـدرة على معالجة البيانات بسرعة، بمجـرد وجودها، حيث إن الطريقة التقليديـة لجمع ومعالجة البيانات ربما لا تؤدي العمل المطلـوب. ولذلك؛ فإنه في العديد من الحالات، ينبغي تحليل البيانات بمجرد الحصول عليها للاستفادة من أكبر قدر ممكن من قيمتها (وهذا ما يسمى بتحليلات التيار، والتي سوف نتناولها لاحقًا في هذا الفصل).
- التحكـم في البيانـات: القدرة على حفظ الأمـن، والسرية، وحقوق الملكية، وجـودة البيانات الضخمة. وينبغي أن تتناسـب قدرات ممارسـة التحكم مع حجم تغيير البيانات، وتنوعها (في التنسيق والمصدر)، وسرعتها.
- توافر المهارات: يتم استغلال البيانات الضخمة من خلال استخدام أدوات جديدة كما يتم البحث عنها بطرق مختلفة. وهناك نقص في الأشخاص (الذين يطلق عليهم غالبًا علماء البيانات) ذوي المهارات اللازمة للقيام بهذه المهمة.
- تكلفة الحل: هناك قدرٌ كبيرٌ من التجارب والاكتشافات التي تجري لتحديد أنماط هذه المسألة والأفكار التي تتحول إلى قيمة، وذلك بسـبب ما قدمته البيانات الضخمة من تحسينات الأعمال الممكنـة. ولضمان تحقيق عائد إيجابي على الاسـتثمار في مشروع البيانـات الضخمة، فإنه من الضروري تقليل تكلفة الحلول المستخدمة لإيجاد تلك القيمة.

وكـما أن التحديات حقيقية، فإن عرض قيمة تحليل البيانات الضخمة يكون أيضًا كذلك. وأي شيء يمكنـك القيام بـه باعتبارك قائد لتحليل الأعمال من أجل المسـاعدة في إثبات قيمـة مصادر البيانات الجديدة بالنسـبة إلى الأعمال، سـوف ينقل مؤسسـتك إلى ما هو أكثر من مجرد تجريب واستكشاف البيانات الضخمة في تكييفها وتبنِّيها باعتبارها شيئًا مختلفًا. ولا يوجد شيء خاطئ مع الاستكشاف، غير أن القيمة تأتي في نهاية المطاف من وضع تلك الأفكار موضع التنفيذ.

مشاكل الأعمال التي تم علاجها من خلال تحليلات البيانات الضخمة:

تتمثـل أهم مشـاكل الأعمال التي تتـم معالجتها عن طريق البيانـات الضخمة بوجهٍ عام في كفاءة العمليات وتخفيض التكاليف، بالإضافة إلى تعزيز تجربة العملاء، غير أنه عندما يتم النظر إليهـا من منظور قطاع الصناعة فـإن هناك أولويات مختلفة تظهر. وربما تكون كفاءة العمليات وتخفيض التكلفة مصنفة من ضمن المشكلات ذات المستوى الأعلى والتي يمكن معالجتها باستخدام تحليلات البيانات الضخمة لقطاعات التصنيع، والحكومة، والطاقة والمرافق، والاتصالات والإعلام، والنقل والرعاية الصحية. كما أن تعزيز تجربة العملاء قد تكون على رأس قائمة المشـكلات التي تعالجهـا شركات التأمين. وبالنسـبة للشركات في القطـاع المصرفي والتعليم فإنـه عادةً ما تكون إدارة المخاطر على رأس القائمة. وفيما يلي قائمة جزئية بالمشـاكل التي يمكن معالجتها باستخدام تحليلات البيانات الضخمة:

- كفاءة العملية وخفض التكلفة.
- إدارة العلامة التجارية.
- تعظيم العائد، والبيع المتقاطع، والبيع من خلال وسطاء.
- تجربة العملاء المحسنة.
- تحديد هوية تحول العملاء، والمحافظة على العملاء.
- خدمة العملاء المحسنة.
- تحديد المنتجات الجديدة وفرص السوق.
- إدارة المخاطر.
- الامتثال للوائح التنظيمية.
- إمكانات الأمان المحسنة.

وتـضرب الحالة العمليـة (٧-٢) مثالًا ممتازًا للصناعة المصرفية، حيث يتـم تكامل المصادر المختلفة للبيانات مع البنية الأساسية للبيانات الضخمة من أجل التوصل إلى مصدر واحد للحقيقة.

حالة عملية ٧-٢

بنـك توب فايف إنفسـتمنت (Top Five Investment) يحقـق التوصل إلى مصدر واحد للحقيقة

يعتبر فريق مشـتقات البنك الذي يحظى بالاحترام الفائق مسؤولًا عن أكثر من ثلث إجمالي تجارة المشـتقات في العالم. كما أن ممارسـة هذا الفريق للمشـتقات لها بصمة

عالمية في مقابل الفِرَق التي تدعم الائتمان، ومعدلات الفائدة، ومشتقات الأسهم في كل منطقــة من مناطق العالم. وقد حصل البنك على العديد من الجوائز في مجال الصناعة والمُعتَرف بها لابتكارات منتجاتها.

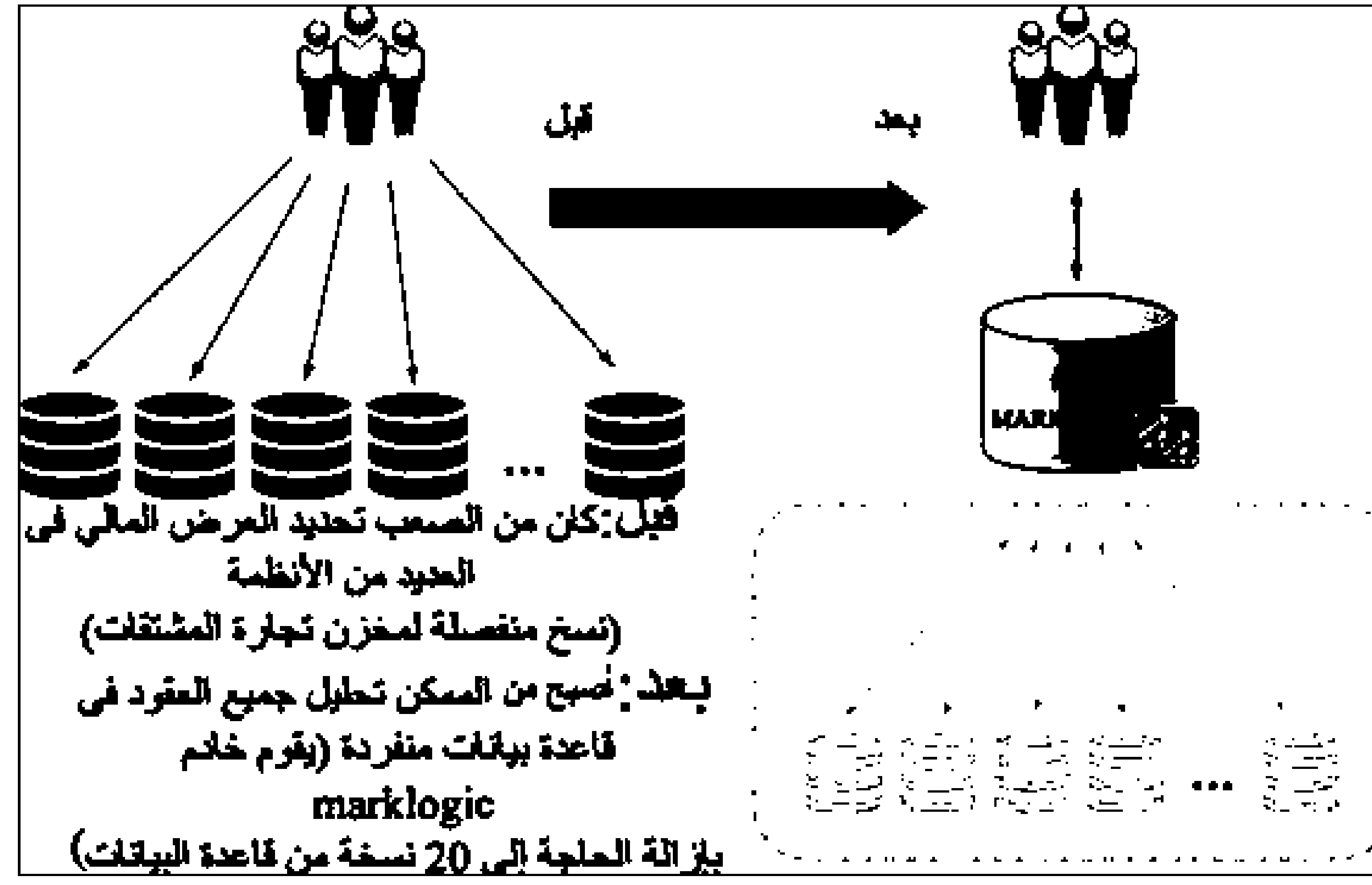
التحدي:

وقــد أدركت إدارة البنك بتعرضها للمشــتقات المهمة، أهميــة وجود رؤية عالمية في الوقــت المناســب لمواقعها. وقد تألف النظــام الحالي، والذي يســتند على قاعدة بيانات علائقية، من العديد من المنشــآت حول العالم. وبسبب التوسعات التدريجية مــن أجل اســتيعاب تزايد أنواع حجم البيانات، فإن النظــام القديم لم يكن سريعًا بما يكفي ليلبي احتياجــات ومتطلبات العمل المتزايدة. حيث لم يتمكن هذا النظام من تقديم تنبيهات في الوقت المناسب لإدارة السوق والمناصب الائتمانية المناظرة في الإطار الزمني المرغوب فيه.

الحل:

قــام البنك ببناء متجر للمشــتقات التجاريــة بناءً عــلى MarkLogic (مورد حلول تحليــلات بيانــات ضخمة)، لتحل محل التقنيات الحالية. وقــد تمكن البنك من معرفة وضعه في الســوق ووضعه الائتماني في الوقت المناسب، وذلك من خلال استبدال خوادم معالجــة الدفعــات المختلفة العشرين بمخــزن تجاري واحد جاهز للتشــغيل، مما وفر القــدرة على التصرف بسرعة للحد من المخاطر. وقد ســمحت دقــة واكتمال البيانات للبنــك وجهاز التنظيم الخاص به، بالاعتماد بشــكل واثق عــلى المقاييس ونتائج اختبار الضغط التي يقدمها.

وقــد اشــتملت عملية الاختيار على ترقيــة كل من تقنيــة Oracle وتقنية Sybase الحاليتــين، بالإضافة إلى الوفاء بجميع المتطلبات التنظيميــة الجديدة والذي يُعد عاملًا رئيسًا في اتخاذ القرار فقد كان البنك يتطلع إلى زيادة استثماراته إلى أقصى حد. وقد كان الخيــار واضحًا بعد التحقيق الدقيق الذي أجراه البنك، فلم يتمكن أي من الخوادم من تلبية كلا الحاجتين فضلًا عن توفير أداء أفضل، وقابلية التوسع، وتطوير أسرع للمتطلبات المســتقبلية والتنفيذية، وتكلفة إجمالية أقل لحقوق الملكية، سوى MarkLogic فقط والذي اســتطاع أن يفي بكل ذلك. ويوضح الشكل (٧-٥) التحويل من الأنظمة المُجزأة القديمة إلى النظام الموحد الجديد.



Sources: MarkLogic.

شكل ٧-٥: الانتقال من عدة أنظمة قديمة إلى نظام جديد موحد

النتائج:

وقع الاختيار على MarkLogic نظرًا لما يوفره من تحديث لجزء من الثانية، بالإضافة إلى أوقات الاستجابة للتحليل اللازمة لإدارة سجل تجارة المشتقات بشكل فعال والذي يمثل ما يقرب من ثلث السوق العالمي، في حين أن النظم الحالية لن تقدم ذلك. والآن يتم تجميع البيانات التجارية بدقة من خلال ملف مشتقات البنك بالكامل، مما يسمح لأصحاب المصلحة بإدارة المخاطر بمعرفة ملف المخاطر الحقيقي للمؤسسة، لإجراء التحليلات التنبؤية باستخدام بيانات دقيقة، ولاعتماد نظرة تطلعية. ولا يقتصر الأمر على توفير مئات آلاف الدولارات من تكاليف التقنية كل عام، بل إن البنك لا يحتاج إلى إضافة موارد للوفاء بالمطالب المتزايدة للهيئات التنظيمية من أجل مزيد من الشفافية وتكرار اختبار الضغط. وفيما يلي أهم العناصر:

- تعمل ميزة التنبيه على إبقاء المستخدمين على دراية فورية بتغييرات السوق والتغييرات الائتمانية لنظرائهم حتى يتمكنوا من اتخاذ الإجراءات المناسبة.

- يتم تخزين المشتقات وتداولها في نظام MarkLogic فردي ولا يتطلب أي توقف للصيانة، مما يعطيه ميزة تنافسية كبيرة.

- يمكن إجراء تغييرات معقدة خلال ساعات في حين أن الخوادم الأخرى تستغرق أيامًا وأسابيع بل وقد تصل إلى أشهر.

- أدى استبدال Oracle وSybase إلى خفض تكاليف العمليات بشكل كبير حيث أنه يوفر: نظام واحد مقابل ٢٠ نظامًا، ومسؤول قاعدة بيانات واحد بدلًا من ١٠، بالإضافة إلى انخفاض تكاليف كل عملية تداول.

الخطوات التالية:

أدى النجــاح في تنفيذ النظــام الجديد وأدائه إلى فحص البنــك لمناطق أخرى حيث يمكنــه أن يســتخلص قيمة أكبر من بياناتــه الضخمة المُهيكلة، أو غــير المُهيكلة، و / أو المتسلســلة. وهناك تطبيقان قيد المناقشة النشــطة. حيث ترى أعمال أبحاث الأسهم الخاصة بهم فرصة لزيادة الإيرادات بشكل كبير من خلال نظام أساسي يوفر الأبحاث في الوقت المناســب، كما يوفر إعادة التهيئة، فضلًا عن المحتوى. كما يرى البنك أيضًا قوة مركزيــة بيانات العملاء في تحســين الإعداد، وزيادة فرص البيــع المتقاطع، ودعم اعرف متطلبات عميلك.

أسئلة للمناقشة:

١- كيف يمكن الاستفادة من البيانات الضخمة في البنوك التجارية على نطاق واسع؟

٢- كيف ساعدت البنية الأساسية لـ MarkLogic في تسهيل الاستفادة من البيانات الضخمة؟

٣- ما هي التحديات، والحل المقترح، والنتائج التي تم التوصل إليها؟

investment bank achieves single ٥ Top .(٢٠١٢) .Sources: MarkLogic derivativestrading--٥-source of truth. marklogic.com/resources/top .(٢٠١٦ bank-achieves-single-source-of-truth (accessed July

أسئلة للمراجعة على القسم ٧-٣:

١- ما هي تحليلات البيانات الضخمة؟ وكيف تختلف عن التحليلات المنتظمة؟

٢- ما هي عوامل النجاح ذات الأولوية لتحليلات البيانات الضخمة؟

٣- ما هي التحديات الكبيرة التي ينبغي مراعاتها عند اعتماد تنفيذ تحليلات البيانات الضخمة؟

٤- ما هي مشاكل العمل الشائعة التي تم علاجها بواسطة تحليلات البيانات الضخمة؟

٧-٤ تقنيات البيانات الضخمة:

هناك عــدد من التقنيات لمعالجة البيانات الضخمة، غير أن معظمها لها بعض الخصائص المشــتركة (Kelly, 2012). وهذا يعني أنهم يســتفيدون من الأجهزة الســلعية لتمكين أساليب التدريج والتعامل

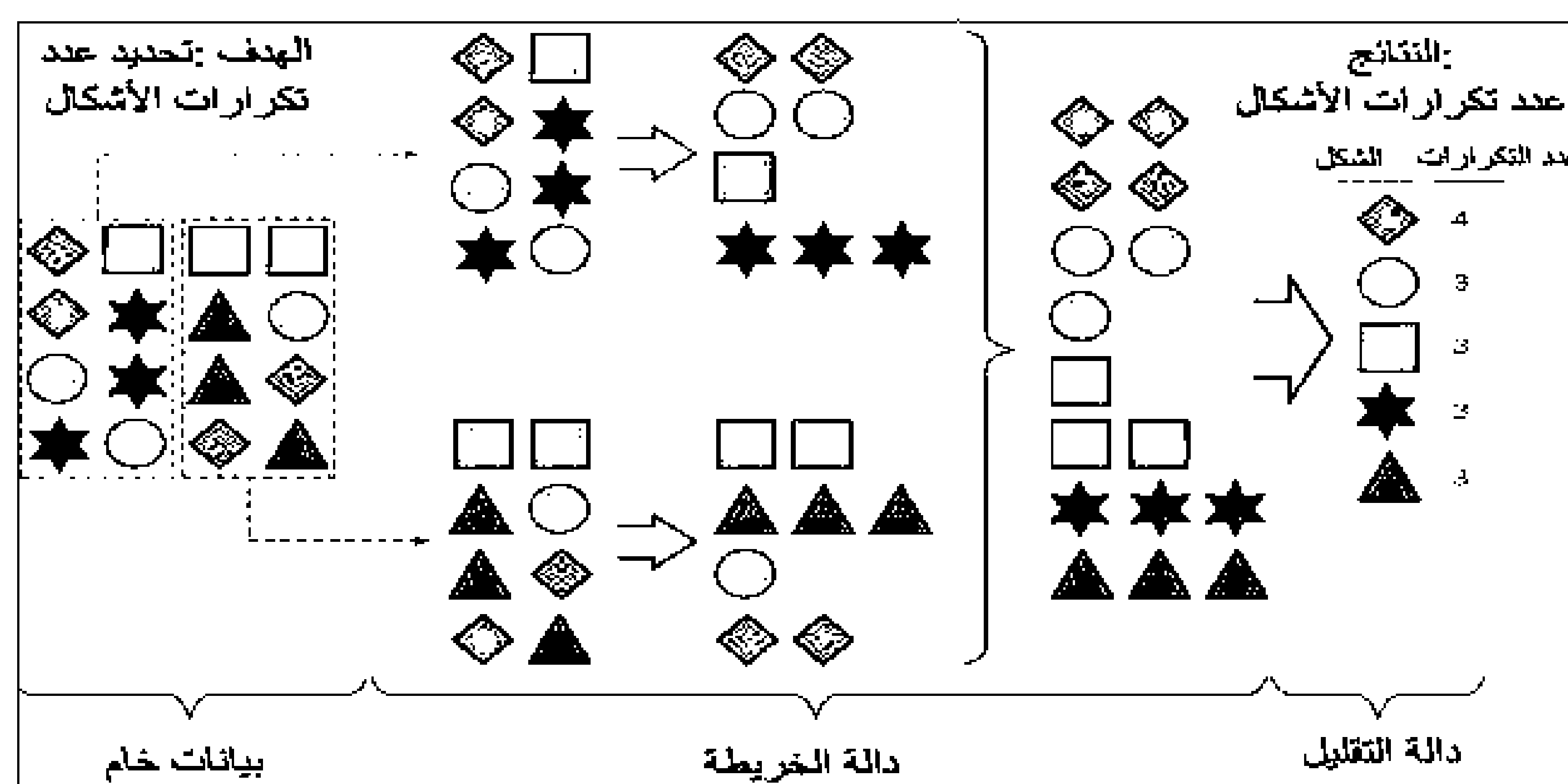
المتوازي؛ واســتخدام قدرات مســتودعات البيانات غير الارتباطية لمعالجة البيانات غير المهيكلة وشــبه المهيكلة؛ وتطبيق تقنية متقدمة للتحليلات والتصوير البصري للبيانات الضخمة من أجل نقل الأفكار إلى المستخدمين النهائيين. وتتمثل تقنيات البيانات الضخمة الثلاثة البارزة في MapReduce، وHadoop، و NoSQL، والتي يُعتقد أنها على الأغلب سوف تحول تحليلات الأعمال وأسواق إدارة البيانات.

:MapReduce

يُعد MapReduce أســلوبًا مفضلًا عن طريق جوجل، والذي يقوم بتوزيع معالجة ملفات البيانات الضخمة متعددة الهيكلة عبر مجموعة كبيرة من الآلات. ويتم تحقيق الأداء العالي عن طريق تقســيم المعالجة إلى وحدات صغيرة من العمل التي يمكن تشغيلها بالتوازي عبر المئات، بل قد تكون عبر الآلاف، من العُقد في مجموعة الآلات تلك. وفيما يلي اقتباس الوثيقة المهمة المتعلقة بـ MapReduce:

> إن MapReduce هــو نموذج برمجة وتنفيذ مترابــط لمعالجة وتوليد مجموعات البيانات الضخمة. حيث تتم موازنة البرامج المكتوبة على هذا النمط الوظيفي تلقائيًا وتنفيذها على مجموعة كبيرة من آلات المادة. الجدير بالذكر أن استخدام مثل هذه الأنظمة المتوازية والموزعة يســمح للمبرمجين الذين ليس لديهم أي خبرة بالاستفادة بسهولة من موارد نظام التوزيع الكبير. (Dean & Ghemawat, 2004).

وتتمثل النقطة الأساســية التي ينبغي ملاحظتها من هــذا الاقتباس في أن MapReduce يُعد نموذج برمجة، وليس لغة برمجة، وهذا يعني أنه مصمم ليتم استخدامه من قِبل المبرمجين، وليس مســتخدمي الأعمال. ولوصف كيفية عمل MapReduce بأســهل طريقة، سنقوم بضرب مثال. انظر عداد المربعات الملونة ColoredSquare في الشكل (٧-٦).



شكل ٧-٦: تصوير بياني لعملية Mapreduce

لاحــظ في هذا الشــكل أن المدخــلات إلى عملية MapReduce هي عبــارة عن مجموعة من المربعــات الملونة. والهدف هو حســاب عــدد مربعات كل لون. وفي هذا المثــال يكون المُبرمج مســؤولًا عن ترميز الخريطة والتقليص من البرامج؛ وما يتبقى من المعالجة يتم التعامل معها من خلال نظام البرمجيات الذي ينفذ نموذج برمجة MapReduce.

حيــث يقوم نظــام MapReduce أولًا بقراءة ملف المدخلات وتقســيمه إلى عدة أجزاء. وفي هذا المثال، هناك نوعان من التقســيمات، بينما على أرض الواقع، عادةً ما يكون عدد التقســيمات أعلى من ذلــك بكثير، ومــن ثَم تتم معالجة هذه التقســيمات عن طريق برامج map المتعــددة والتي تعمل بالتــوازي على عُقد المجموعــة. وفي هذه الحالة يكون دور كل برنامج هو تجميع وتقســيم البيانات حســب اللــون، ثم يقوم نظام MapReduce بأخذ مخرجات كل برنامــج ودمج (خلط / فصل) نتائج المدخلات إلى برنامج التقليص، والذي يقوم بحساب مجموع عدد المربعات لكل لون. وفي هذا المثال، يتم اســتخدام نســخة واحدة فقط من برنامج التقليص، غير أنه من الناحية العملية قد يكون هناك المزيد من النُسخ. ويمكن للمبرمجين توفير برنامج الخلط / الفصل الخاص بهم، من أجل تحسين الأداء كما يمكنهم أيضًا نشر أداة الدمج التي تجمع بين ملفات مخرجات الخريطة المحلية؛ وذلك لتقليل عدد ملفات المخرجات التي يجب الوصول إليها عن بُعد عبر المجموعة عن طريق خطوة الخلط / الفصل.

لماذا يُستَخدَم MapReduce؟

يقوم MapReduce بمساعدة المنظمات في معالجة وتحليل كميات كبيرة من البيانات متعددة المراحل. وتنطوي أمثلة التطبيق على الفهرســة والبحث، والتحليــلات البيانية، وتحليلات النص، وتعلم الآلة، وتحويل البيانات وهكذا. وغالبًا ما تكون هذه الأنواع من التطبيقات صعبة التنفيذ باستخدام SQL القياسية المستخدمة بواسطة DBMSs العلائقية.

إن طبيعة MapReduce الإجرائية تجعله ســهل الفهم بواســطة المبرمجين ذوي المهارة. كما أنــه أيضًا له ميزة أخرى وهي أن المطورين ليســوا مُلزَمين بالاهتمام بتنفيذ الحوســبة المتوازية، حيث يتعامل النظام مع هذا الأمر بشفافية. وعلى الرغم من أن MapReduce تم تصميمه من أجل المبرمجين، إلا أن غير المبرمجين يمكنهم أيضًا اســتغلال قيمة كل من تطبيقات MapReduce ومكتبات الوظائف. وتتوفر مكتبات MapReduce سواء المدفوعة أو المجانية والتي توفر نطاقًا واســعًا من القدرات التحليلية. فعلى سبيل المثال، نجد أن مكتبة Apache Mahout هي مكتبة تعلم آلي مجانية لـ "خوارزميات التجميع والتصنيف والترشــيح التعاوني المستند إلى الدفعة" التي يتم تنفيذها باستخدام MapReduce.

:Hadoop



Source: Hadoop.

يعتـبر Hadoop إطـار عمل مجاني لمعالجـة، وتخزين، وتحليل كميات ضخمـة من البيانات المتفرقة وغير المهيكلة. وقد تم إنشاء Hadoop في البداية من قِبل Doug Cutting على Yahoo، وقد تم اسـتخلاص فكرة تقنية Hadoop مـن MapReduce، والذي هو عبارة عن دالة محددة بواسطة المستخدم ومطورة بواسطة جوجل في بدايات الألفية الثانية من أجل فهرسة الويب. وقد تم تصميمه للتعامل مع البيتابايتس والإكسـابايتس الخاصة بالبيانات التي تم توزيعها عبر عقد متعددة في نفس الوقت. وتعمل مجموعات Hadoop على أجهزة مواد غير مُكلِفَة بحيث يمكن توسيع نطاق المشرعات دون اللجوء للبنك. وقد أصبح Hadoop الآن مشروعًا لتأسيس برمجيات أباتـشي Apache Software Foundation، حيث يعمل المئات من المسـاهمين باسـتمرار على تحسـين التقنية الأساسية. وتتمثل الفكرة الأساسية في: قيام Hadoop بتجزئة البيانات الضخمة إلى أجـزاء متعددة بحيث يمكن معالجـة كل جزء وتحليله في نفس الوقت، بدلًا من كثرة التخبط في مجموعة ضخمة من البيانات باستخدام جهاز واحد.

كيف يعمل Hadoop؟

يقـوم العميل بالوصول إلى البيانات غير المهيكلة وشـبه المهيكلة مـن المصادر بما في ذلك ملفات السجل، وإشعارات وسائل التواصل الاجتماعي، ومخازن البيانات الداخلية، ثم يقوم بتقطيع البيانات إلى «أجـزاء»، وهذه الأجزاء يتـم تنزيلها بعد ذلك في نظام ملفات يتألـف من عُقد متعددة تعمل على أجهزة مادية. وأما مخزن الملفات الافتراضي في Hadoop فهو نظام ملفات Hadoop المُقَسـمة Hadoop Distributed File System أو HDFS. ومـن الجديـر بالذكـر أن أنظمة الملفات مثل نظـام HDFS هي أنظمة بارعة في تخزين كميات ضخمة من البيانات غير المهيكلة وشـبه المهيكلة نظـرًا لأنهـا لا تتطلب تنظيم البيانات في الصفوف والأعمدة العلائقية. حيث يتم نسـخ كل «جزء» عدة مرات ثم يتم تنزيله في نظام الملفات بحيث إذا سقطت عُقدة، يكون هناك عُقدة أخرى لديها نسخة من البيانات الموجودة على العُقدة التي سقطت. وتعمل عُقدة الاسم كمسِهل، فهي ترجع إلى معلومات العميل مثل العُقد المتاحة، حيث تتواجد بيانات معينة في المجموعة، والعُقد التي سقطت.

وبمجرد تنزيل البيانات في المجموعة، فإنها تكون جاهزة للتحليل من خلال إطار عمل MapReduce. ويقـوم العميل باختيار وظيفـة "Map" - وعادةً ما تكون عبارة عن اسـتعلام مكتوب بلغة Java -

إلى إحـــدى عُقـــد المجموعة والمعروفة على أنها متعقب الوظائف، حيث يشـــير متعقب الوظائف إلى عُقدة الاســـم لتحديد البيانات التي يحتاج إليها للوصول إلى إكمال هذه الوظيفة والمكان الموجود في المجموعة التي توجد فيها البيانات. وعند تحديد ذلك يقوم مُتَعَقب الوظائف بإرســـال الاستعلام إلى العُقـــد ذات الصلة. وبدلًا من إعادة إدخال جميـــع البيانات إلى موقع مركزي من أجل المعالجة، فإن المعالجة تحدث في كل عُقدة في وقت واحد أو بشكل متوازٍ. وهذه سمة أساسية خاصة بـ Hadoop.

وعندمـــا تنتهـــي كل عقدة من معالجة مهمتهـــا، فإنها تقوم بتخزين النتائـــج. ويقوم العميل بوظيفة "التقليص" من خلال متعقب الوظائف حيث يتم تجميع نتائج مرحلة الخريطة المُخزنة داخليًا على العُقد الفردية لتحديد «الإجابة» على الاستعلام الأصلي، وعندئذٍ يتم تنزيلها على عُقدة أخـــرى في المجموعة. وبوصول العميل إلى هذه النتائـــج، والتي يمكن تنزيلها بعد ذلك في واحدة من العديد من البيئات التحليلية من أجل تحليلها. تكون وظيفة MapReduce قد اكتملت.

وبمجـــرد أن تكتمل مرحلة MapReduce، فإن البيانات المعالجـــة تصبح جاهزة لإجراء المزيد من التحليلات بواســـطة علماء البيانات وغيرهم مـــن أصحاب المهارات المتقدمة لتحليل البيانات. حيث يســـتطيع علماء البيانات التحكم فيها وتحليلها مســـتخدمين أي عدد من الأدوات لأي عدد من المرات، بما في ذلك البحث عن الأفكار والأنماط الخفية، أو اســـتخدامها كأساس لبناء تطبيقات تحليليـــة تواجه المســـتخدم. كما يمكن أيضًـــا تصميم البيانات ونقلها مـــن مجموعات Hadoop إلى قواعد البيانات العلائقية، ومســـتودعات البيانات، وأنظمة تقنية المعلومات التقليدية الأخرى لإجراء مزيد من التحليل و / أو لدعم معالجة المعاملات.

مكونات Hadoop الفنية:

تتكون "حزمة" Hadoop من عدد من المكونات، والتي تشمل:

- نظـــام ملفات Hadoop المقســـمة (HDFS): وهي طبقة التخزيـــن الافتراضية في أي مجموعة محددة من مجموعات Hadoop.
- عُقدة الاسم: وهي العقدة في مجموعة Hadoop التي توفر معلومات العميل فيما يتعلق بمكان مستودعات البيانات الخاصة بالمجموعة وما يتعلق بحالة ما إذا فشلت إحدى العقد.
- العُقدة الثانوية: وهي نســـخة احتياطية من اسم العُقدة، ففيها يتم إجراء نَسْخ متماثل بشكل دوري ومستودعات البيانات من عقدة الأسماء في حالة فشلها.
- مُتعقب الوظائف: وهو عبارة عن العُقدة في مجموعة Hadoop والتي تقوم بافتتاح وتنســـيق وظائف MapReduce أو بمعالجة البيانات.

- العُقد التابعة: وهي العمق الخاص بأي مجموعة Hadoop، وتقوم العُقد التابعة بمستودعات البيانات واتخاذ التوجيه لمعالجتها من متعقب الوظائف.

وبالإضافة إلى هذه المكونات، فإن النظام البيئي Hadoop يتكون من عدد من المكونات الفرعية التكميلية. فهناك مخازن البيانات NoSQL مثل Cassandra وHBase والتي تُستخدم أيضًا لتخزين نتائج وظائف MapReduce في Hadoop. وإلى جانب لغة Java فإن هناك لغة Pig، وهي لغة متاحة مجانًا ومصممة خصيصًا من أجل Hadoop، وتُستخدم لكتابة بعض وظائف MapReduce والوظائف الأخرى لـ Hadoop. كما أن هناك Hive وهو عبارة عن مستودع بيانات مجاني يسمح بالنمذجة التحليلية داخل Hadoop وقد تم تطويره في الأصل بواسطة Facebook. وفيما يلي مكونات Hadoop الفرعية الأكثر شيوعًا.

Hive: HIVE هو عبارة عن إطار عمل لمستودعات البيانات القائم على Hadoop وقد تم تطويره من قِبل Facebook، وهو يسمح للمستخدمين بكتابة الاستعلامات بلغة مثل SQL وتسمى HiveQL، والتي يتم تحويلها بعد ذلك إلى MapReduce، مما يسمح لمبرمجي SQL الذين ليس لديهم خبرة بـMapReduce باستخدام المستودع وجعله أسهل لدمجه مع ذكاء الأعمال وأدوات التمثيل المرئي مثل: Microstrategy، وTableau، وRevolutions Analytics، وغير ذلك.

PIG: وهي عبارة عن لغة استعلام قائمة على Hadoop تم تطويرها بواسطة Yahoo وهي سهلة التعلم نسبيًا كما أنها بارعة في امتدادات البيانات الطويلة جدًا والعميقة للغاية (حدود SQL).

HBASE: إن HBASE هي قاعدة بيانات غير علائقية تتيح عمليات بحث سريع في برنامج Hadoop في وقت قليل، كما أنها تضيف إمكانيات المعاملات إلى Hadoop، مما يسمح للمستخدمين بإجراء التحديثات، وعمليات الحذف والإضافة. ويستخدم كل من eBay و Facebook قاعدة بيانات HBASE بكثرة.

FLUME: وهو إطار عمل لتغذية Hadoop بالبيانات. حيث يتم ملء كل مكان في البنية الأساسية لتقنية المعلومات تقنية المعلومات بالعملاء - داخل خوادم الويب، وخوادم التطبيقات، وأجهزة الجوال - على سبيل المثال: لجمع البيانات ودمجها في Hadoop.

OOZIE: وهو نظام معالجة سير العمل الذي يسمح للمستخدمين بتحديد سلسلة من الوظائف المكتوبة بلغات متعددة مثل MapReduce، وPig، وHiveومن ثم ربطها ببعضها البعض بذكاء. حيث يتيح Oozie للمستخدمين تحديد أشياء، مثل: أنه لا يتم طلب استعلام خاص إلا بعد انتهاء الوظائف السابقة المحددة التي يعتمد عليها من أجل البيانات.

AMBARI: وهـي مجموعة مـن الأدوات التي تعتمد على الويـب لنشر مجموعات Apache Hadoop، وإدارتهـا، ومراقبتهـا. ويقـوم بقيـادة تطويرهـا مهندسـون مـن شركـة هورتنـورك (Hortonworks)، والتي تشتمل على Ambari في منصة بيانات هورتنورك الخاصة بها.

AVRO: وهو نظام تسلسـل البيانات الذي يسـمح بترميز مخطط ملفات Hadoop، وهو بارع في تحليل البيانات والقيام بإزالة استدعاءات الإجراءات.

MAHOUT: وهي مكتبة للتنقيب في البيانات، وهي تستخدم أكثر خوارزميات جمع البيانات انتشارًا للقيام بالتجميع، واختبار الانحدار، والنمذجة الإحصائية، وتنفيذ كل ذلك باستخدام نموذج MapReduce.

SQOOP: وهـو أداة اتصـال لنقل البيانات من مخازن بيانـات غير Hadoop - مثل قواعد البيانات العلائقية ومستودعات البيانات - إلى Hadoop، وهو يسمح للمستخدمين بتحديد موقع الهدف داخل Hadoop فضلًا عن إرشاد Sqoop لنقل البيانات من Oracle، أو تيراداتا، أو غيرها من قواعد البيانات العلائقية إلى الهدف.

HCATALOG: وهـي عبـارة عن إدارة مركزيـة للبيانات الوصفية metadata ومشـاركة الخدمـات لـــ Apache Hadoop، وهي تسـمح برؤية موحدة لجميع البيانـات في مجموعات Hadoop كـما تتيح أدوات متنوعة تتضمن كلًا من Pig وHive، لمعالجة أي عناصر بيانات دون الحاجة إلى معرفة مادية بمكان مستودعات البيانات في المجموعات.

Hadoop: الإيجابيات والسلبيات:

تتمثل الميزة الأساسـية لـ Hadoop في أنه يسـمح للمشـاريع بمعالجة وتحليل كميات كبيرة مـن البيانات غير المهيكلة وشـبه المهيكلة، والتي يتعذر الوصول إليهـا حتى الآن، بطريقة فعالة مـن حيث التكلفـة والوقت. ونظرًا لأن مجموعـات Hadoop يمكنها الوصـول إلى البيتابايتس وحتى الإكسابايتس من البيانات، فإن المشاريع لم تعد بحاجة إلى الاعتماد على مجموعات عينات البيانـات ولكن يمكنها معالجة وتحليل كل البيانـات ذات الصلة. ويمكن لعلماء البيانات تطبيق تدخل الحلقات التكرارية، حيث إن تنقيح واختبار الاسـتعلامات باستمرار يؤدي إلى الكشف عن أفكار لم تكن معروفة مسـبقًا. أيضًا من مميزات Hadoop أن البدء في اسـتخدامه لا يحتاج إلى تكلفة عالية، بالإضافة إلى أن المطورين يمكنهم تنزيل تقسيمات Apache Hadoop مجانًا والبدء بتجريب Hadoop في أقل من يوم واحد.

أما السـلبيات الخاصة بـ Hadoop ومكوناته المتعـددة فهي تتمثل في أنه لم يصل إلى مرحلة النضـج كـما أنه لازال في تطور حتى الآن. وكما هو الحال مع أي تقنية ناشـئة وغير ناضجة، فإن

تنفيـذ وإدارة مجموعات Hadoop بالإضافة إلى إجراء تحليلات متقدمة على كميات ضخمة من البيانـات غير المهيكلة يتطلـب خبرةً ومهارةً وتدريبًا عاليًا. ولسـوء الحظ، فإنه في الوقت الحالي هنـاك ندرة في مطوري Hadoop وعلماء البيانات، مما يجعل الحفاظ على مجموعات Hadoop المعقدة والاستفادة منها أمرًا ليس عمليًا. وبالإضافة إلى ذلك، فإنه مع تحسين مكونات Hadoop العديـدة في المجتمع، وتكوين المكونات الجديدة، فإن هنـاك خطورة من التفرع، وذلك كما هو الوضـع مع أي تقنيـة / منهج مجاني غير ناضج. وفي النهاية، فـإن Hadoop هو عبارة عن إطار موجه بالدفعات، بمعنى أنه لا يدعم معالجة وتحليل البيانات بشكل فوري.

والمفاجأة السـارة أن بعض ألمع العقول في مجال تقنية المعلومات يساهمون في مشروع Apache Hadoop، وفضلًا عن ذلك فإن هناك جيلًا جديدًا من مطوري Hadoop ومن علماء البيانات ممن هم في مقتبل العمر. ونتيجة لذلك؛ تتقدم التقنية بسرعة لتصبح أكثر قوة وأسهل في التنفيذ والإدارة. وأما النظـام البيئي للموردين، فإنه يتمثـل في عمل كل من Hadoop-focused والشركات حديثة الظهور مثـل Cloudera وHortonworks والـشركات الرائـدة في تقنية المعلومات مثـل أي بي إم (IBM)، ومايكروسـوفت (Microsoft)، وتيراداتـا (Teradata)، وأوراكل (Oracle) لتقديم كل من توزيعات Hadoop الجاهزة للشركات، والأدوات، والخدمات التجارية وذلك لجعل نشر التقنيات وإدارتها واقعًا عمليًا للمؤسسة التقليدية. وتعمل الشركات الناشئة الأخرى والتي هي في مقدمة التقنية، على تحسين مخـزون البيانـات NoSQL (ليس فقط SQL) القادر على تقديم أفـكار قريبة فوريًا مع Hadoop. وتقدم رؤية فنية (٧-٢) بعض الحقائق لتوضيح بعض المفاهيم الخاطئة حول Hadoop.

رؤية فنية ٧-٢

توضيح بعض الحقائق الغامضة الخاصة بهادوب

عـلى الرغـم من وجود هادوب والتقنيـات المرتبطة به منذ أكثر من ٥ سـنوات، لازال معظم الناس يؤمنـون ببعض المفاهيم الخاطئة حـول هادوب والتقنيات المرتبطة به مثـل: MapReduce وHive. وسـوف تقدم القائمة التالية والتي تتكون من ١٠ حقائق توضيحًا لماهية هادوب وما يقوم به بالنسـبة إلى ذكاء الأعمال، وكذلك حالات الأعمال والتقنية التي بإمكانها الاستفادة من ذكاء الأعمال ومستودعات البيانات والتحليلات القائمة على هادوب (Russom, 2013).

الحقيقـة # ١: يتألـف هادوب من برامج متعددة. دائمًـا ما نتحدث عن هادوب كما لو أنه برنامج واحد مترابط، في حين أنه في الواقع عبارة عن عائلة من المنتجات والتقنيات المجانية والتي تشرف عليها مؤسسـة برنامج Apache (ASF). وتتوفر بعض منتجات هادوب أيضًـا من خلال توزيعات الموردين؛ وسيتم تقديم المزيد عن ذلك لاحقًا.

وتشمل مكتبة Apache Hadoop (في ذكاء الأعمال حسب الأولوية)، كلًا من HDFS، وMapReduce، وHive، وHbase، وPig، وZookeeper، بالإضافة إلى Flume، وSqoop، وOozie، وHue، وما إلى ذلك. والتـــي يمكنك دمجها بطرق متنوعة، غير أن HDFS وMapReduce (عند دمجهم مع Hbase وHive) قد يشكلان حزمة تقنية مفيدة لكل التطبيقات في ذكاء الأعمال، ومستودعات البيانات، والتحليلات.

الحقيقـــة # ٢: يتوفـــر هـــادوب مجانًا ومفتوح المصدر لكنـــه متاح أيضًا من قِبـــل الموردين، حيث إن مكتبـــة برامـــج Apache Hadoop المجانيـــة تكون متاحة من ASF على apache.org. وبالنســـبة للمستخدمين الذين يرغبون في الحصول على حزمة أكثر جاهزية للمؤسسة، فإن عددًا قليلًا من الموردين الآن يقدمون تقسيمات هادوب التي تتضمن أدواتٍ إدارية إضافية ودعمًا فنيًا.

الحقيقة # ٣: يعد هادوب نظامًا بيئيًا، وليس منتجًا مســـتقلًا؛ إذ يشـــتمل نظام هادوب البيئي على قائمة متزايدة من برامج الموردين التي تتكامل مع أو تنشر تقنيات هادوب، ولن تحتاج لأكثر من دقيقة واحدة على محرك بحثك المفضل لتكتشف ذلك.

الحقيقـــة # ٤: يعتبر HDFS نظـــام ملفات، وليس نظام إدارة قاعدة بيانات (DBMS). ويُعد هادوب نظام ملفات مقســـمة في المقام الأول، ويفتقر إلى القدرات التي يمكننا ربطها بنظام DBMS، مثل الفهرسة، والوصول العشوائي إلى البيانات، والدعم لـ SQL. وهذا جيد، نظرًا لأن نظام HDFS يقوم بأشياء لا تستطيع قواعد البيانات العلائقية (DBMSs) القيام بها.

الحقيقة # ٥: يتشابه Hive مع SQL ولكنه ليس هو SQL القياسي. ويقصر العديد منا نفسه على SQL وذلـــك نظـــرًا لمعرفتنا الجيدة به بالإضافة إلى أن أدواتنا تتطلب ذلك. وبالنســـبة لمن يعرفون لغة SQL فإنهـــم يســـتطيعون تعلم الرمز اليدوي Hive بسرعة، غير أن هذا لا يحل مشـــكلات التوافق مع الأدوات القائمة على SQL.

الحقيقة # ٦: هناك ارتباط بين هادوب وMapReduce غير أن هذا الارتباط لا يقتضي أن يحتاج كل منهـــما للآخر. وقد قام المطورون على جوجـــل بتطوير MapReduce قبل وجود HDFS، وهناك بعض أشكال MapReduce التي تعمل مع مجموعة متنوعة من تقنيات التخزين، بما في ذلك HDFS، وأنظمة الملفات الأخرى، وبعض DBMSs.

الحقيقة # ٧: يوفر MapReduce التحكم في التحليلات، وليس التحليلات في حد ذاتها؛ حيث إن MapReduce هو محرك تنفيذ أغراض عامة والذي يتعامل مع تعقيدات الاتصال بالشـــبكة، والبرمجة المتوازية، واحتمال الخطأ لأي نوع من التطبيقات التي يمكنك من خلالها تقديم رمز، وليس مجرد تحليلات.

الحقيقـــة # ٨: يـــدور هادوب حول تنوع البيانات، وليس فقط حول حجـــم البيانات. فمن الناحية النظرية، يمكن لنظام HDFS إدارة التخزين والوصول إلى أي نوع من البيانات طالما يمكنك وضع البيانات

في ملف ونسـخ هذا الملف إلى HDFS. ونظرًا لأن هذا الأمر يبدو بسـيطًا للغاية، كما أنه صحيح بصفة عامة، فهذا هو ما يجذب العديد من المستخدمين لـ Apache HDFS.

الحقيقة # ٩: يعمل هادوب كمكمل لمستودعات البيانات؛ ونادرًا ما يكون بديلًا عنه. فقد صممت معظم المنظمات مسـتودعات البيانات الخـاص بها من أجل البيانات المهيكلـة، والعلائقية، مما يجعل من الصعب حصر قيمة ذكاء الأعمال من البيانات غير المهيكلة وشـبه المهيكلة. ويقوم هادوب بتكميل مسـتودعات البيانات من خلال التعامـل مع أنواع البيانات متعددة الهياكل والتي لا تسـتطيع معظم مستودعات البيانات (DWs) التعامل معها.

الحقيقة # ١٠: يتيح هادوب أنواعًا عديدة من التحليلات، وليس فقط تحليلات الويب. فعلى الرغم من أن هادوب يقع عليه الكثير من الضغوط حول كيفية اسـتخدام شركات الإنترنت له من أجل تحليل سـجلات الويب وبيانات الويب الأخرى، إلا أنه توجد حالات أخرى يُسـتخدم فيها. فعلى سـبيل المثال، تأمـل البيانات الضخمة التي تأتي من أجهزة الاستشـعار، مثل: الروبوتات في التصنيع، أو RFID في بيع التجزئـة، أو مراقبة الشـبكة في المرافق العامة. كما يمكن للتطبيقـات التحليلية القديمة التي تحتاج إلى عينات كبيرة من البيانات - مثل تجزئة قاعدة العملاء، والكشف عن الغش، وتحليل المخاطر - أن تستفيد مـن البيانات الضخمة الإضافية التي يديرها هادوب. وبالمثـل، فإنه يمكن لبيانات هادوب الإضافية أن توسع نطاق المشاهدات ٣٦٠ درجة من أجل إنشاء عرض أكثر اكتمالًا ودقة.

NoSQL:

ظهـر نمط جديد من قاعـدة البيانات ذات الصلة يسـمى NoSQL (ليس فقـط SQL)، وهو مثل هادوب، حيث يعالج كميات ضخمة من البيانات متعددة الهياكل. ومع ذلك، فإنه في الوقت الذي يعد فيه هادوب بارعًا في دعم التحليلات التاريخية على نطاق واسع، فإن قواعد بيانات NoSQL موجهة في معظمها (على الرغم من أن هناك بعض الاستثناءات المهمة)، لتقديم بيانات منفصلة مخزنة بين كميات ضخمة من البيانات متعددة المراحل للمستخدم النهائي وجعل تطبيقات البيانات الضخمة آلية. وتفتقر هذه الإمكانية بشدة إلى تقنية قواعد البيانات العلائقية، وهذا يجعلها ببساطة لا تستطيع الحفاظ على مستويات أداء التطبيقات المطلوبة في مقياس البيانات الضخمة.

وفي بعض الحالات، يعمل كل من NoSQL وهادوب معًا في نفس الوقت. فعلى سبيل المثال، HBase المذكورة آنفًا، هي قاعدة بيانات NoSQL منتشرة على غرار Google BigTable والتي غالبًا ما يتم نشرها على رأس HDFS، ونظام الملفات المقسـمة Hadoop، من أجل توفير عمليات بحث سريعة في هادوب في وقت قليل. ويتمثل الجانب السلبي لمعظم قواعد بيانات NoSQL اليوم في أنها تتعامل مع ACID

(الانســيابية atomicity، والاتســاق consistency، والانعزال isolation، والمتانة durability) الالتزام بالأداء وقابلية التوســع. كما يفتقر الكثير منها إلى وجود أدوات الإدارة والرصد الناضجة. ويتم التغلب على القصور في هذين الجانبين الســلبيين بواسطة مجموعات NoSQL المجانية ومجموعة من الموردين الذيــن يحاولون تســويق قواعد بيانات NoSQL المختلفة. وتشــتمل قواعد بيانــات NoSQL المتاحة حاليًا عــلى HBase، وCassandra، وMongoDB، وAccumulo، بالإضافة إلى Riak، وCouchDB، و DynamoDB، وغيرها. وتعرض الحالة العملية (٧-٣) استخدام قواعد بيانات NoSQL في eBay. كما توضح الحالة العملية (٧-٤) تطبيقًا للتواصل الاجتماعي حيث تم اســتخدام البنية الأساســية لــهادوب لتجميع مجموعة من الرســائل على تويتر لفهم أي أنواع المســتخدمين الذين يشــاركون في أي نوع من الدعم لمرضى الرعاية الصحية الذين يبحثون عن معلومات حول الأمراض العقلية المزمنة.

حالة عملية ٧-٣

حل البيانات الضخمة الخاص بـ eBay

يُعد eBay هو أكبر ســوق على الإنترنت في العالم، مما يتيح شراء وبيع أي شيء عمليًا. وقــد تأســس هذا الســوق في عام ١٩٩٥، حيث يربــط eBay بين أفــراد مجتمع متنوع وتفاعلي من البائعين والمشــترين، فضلًا عن الــشركات الصغيرة. الجدير بالذكر أن eBay الجماعــي له تأثير مذهل على التجارة الإلكترونية: فقد بلغت القيمة الإجمالية للســلع المباعــة على موقــع eBay نحو ٧٥,٤ مليــار دولار في عام ٢٠١٢. ويقــدم موقع eBay خدمات لأكثر من ١١٢ مليون مستخدم نشط كما يقدم أكثر من ٤٠٠ مليون عنصر للبيع.

التحدي - دعم البيانات على نطاق واسع:

تُعــد قــدرة eBay على تحويل كميــات ضخمة من البيانات التــي تنتجها إلى أفكار مفيدة تُمكِّن عملاءها من اســتخلاصها مباشرة من الصفحات المتكررة، هي أحد مفاتيح نجاحه الاســتثنائية. ولاســتيعاب النمو الضخم في بيانات eBay - فــإن مراكز البيانات التابعة له تقوم بأداء المليارات من القراءات والكتابات كل يوم - وبسبب زيادة الطلب على معالجة البيانات بسرعات عالية، فقد احتاج موقع eBay إلى حل ليس له اختناقات مثيلة، بالإضافة إلى أنه قابل للتوســع، وقيود المعامــلات المرتبطة بنهج قواعد البيانات العلائقية الشائعة. أيضًا، فقد كانت الشركة بحاجة إلى إجراء تحليل سريع على تشكيلة واسعة من البيانات المهيكلة وغير المهيكلة التي تم التقاطها.

الحل - البيانات الفورية المتكاملة والتحليلات:

إن متطلبــات البيانات الضخمة جعلت eBay ينتقل إلى تقنيات NoSQL، وتحديدًا Apache Cassandra وDataStax Enterprise. وقد انجذب eBay أيضًا إلى تحليلات

Apache Hadoop المدمجة مع DataStax Enterprise، إلى جانب Cassandra التي تتميــز بياناتها بالفئة عاليــة السرعة. وينطوي الحل على بنيــة تدريجية تجعل eBay قادرًا على نشر مجموعات من DataStax Enterprise المتعددة عبر عدة مراكز بيانات مختلفة باســتخدام الأجهزة السلعية. والنتيجة النهائية هي أن موقع eBay أصبح الآن قادرًا على معالجة كميات هائلة من البيانات بكفاءة عالية من حيث التكلفة، وبسرعات عالية جدًا، عند سرعات عالية للغاية، كما أنه أصبح قادرًا على تحقيق أكثر مما كان قادرًا على تحقيقه مع نظام الملكية الأعلى تكلفة والذي كان يســتخدمه. ويقوم eBay حاليًا بإدارة جزء كبير من احتياجات مركز البيانات - أكثر من ٢٥٠ تيرابايت من المخزون - في مجموعات كل من Apache Cassandra وDataStax Enterprise.

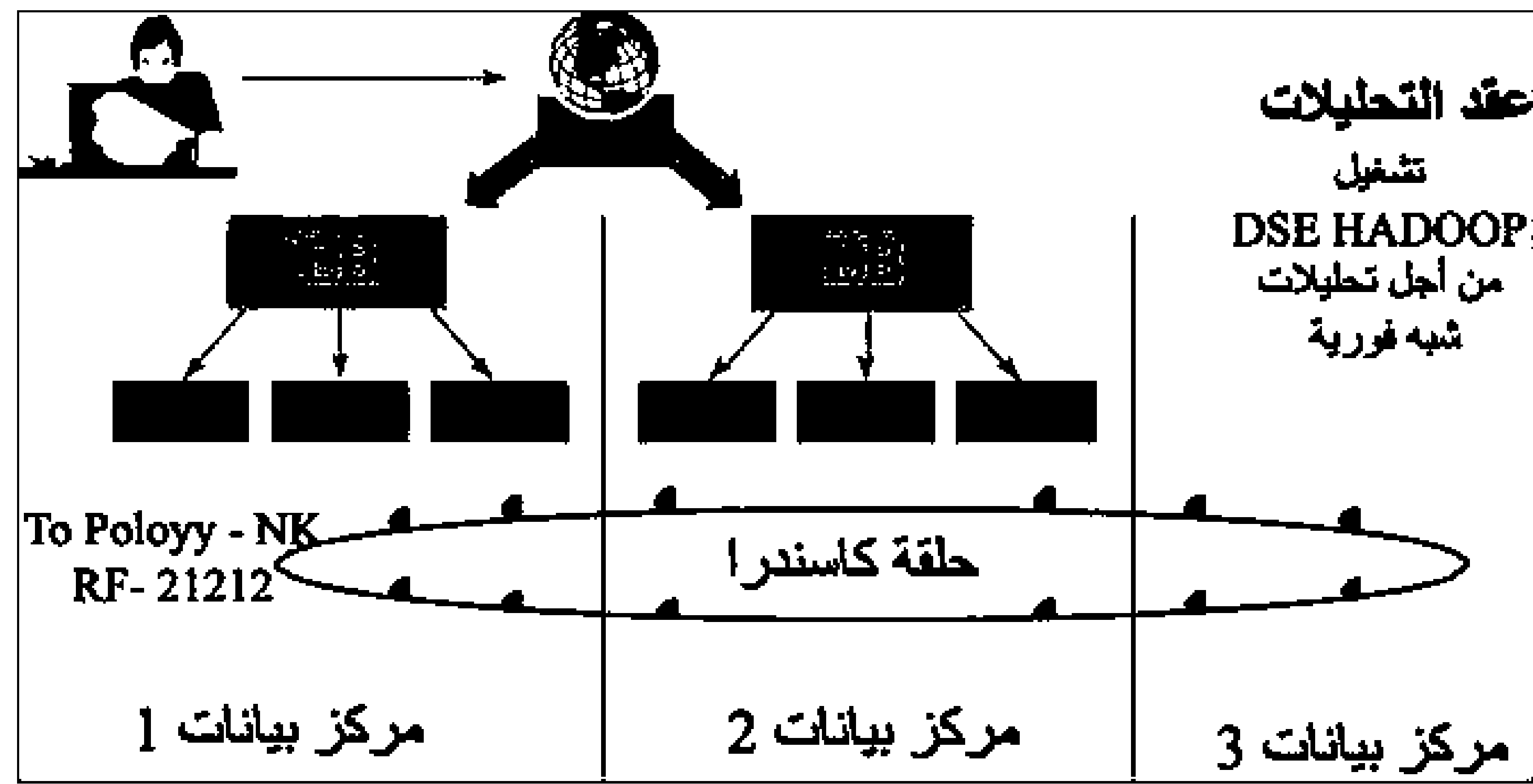
وتشــمل العنــاصر الفنية الإضافية التي لعبت دورًا في قــرار eBay لنشر DataStax Enterprise على نطاق واســع، كلًا من قابليــة التطوير الخطية للحل، والأمل الكبير في النجاح دون نقطة فشل واحدة، وأداء كتابة متميز.

التعامل مع حالات الاستخدام المتنوعة:

يقــوم eBay باســتخدام DataStax Enterprise للعديــد مــن حالات الاســتخدام المختلفــة. وتوضح الأمثلــة التالية بعض الطرق التي تســتطيع الشركة من خلالها تلبية احتياجات بياناتها الضخمة من خلال قدرات معالجة البيانات والتحليلات السريعة التي يوفرها الحل. وبطبيعة الحال، فإن eBay تجرب كميات ضخمة من ســير الكتابة، والتي يتولى تنفيذها تطبيق Cassandra عند معالجــة DataStax Enterprise بطريقة أكثر كفــاءة من أي حــل يقدمــه RDBMS أو NoSQL. وتشــاهد eBay حاليًا أكثر من ٦ مليارات مــن الكتابات يوميًا عبر مجموعات Cassandra المتعددة وأكثر من ٥ مليارات من القراءات (معظمها غير متصل) يوميًا.

وتنطوي حالة الاســتخدام الواحدة التي يدعمها DataStax Enterprise على قياس بيانات eBay الاجتماعية التي يعرضها على صفحات المنتج الخاصة به. ويقوم تقســيم Cassandra في DataStax Enterprise بتخزيــن جميــع المعلومــات اللازمة لتقديم بيانــات لـ "يحب"، و"يفضــل"، و"يريد" على صفحات منتجات eBay. كما يوفر نفس البيانــات الخاصة بصفحــة "تفضيلاتك" على موقــع eBay والتي تحتــوي على جميع العناصر التي يحبها المســتخدم أو يمتلكها أو يرغــب في شرائها، مع عرض Cassandra لصفحــة "تفضيلاتك" كاملة. وتوفر eBay هذه البيانــات من خلال خاصية العدادات القابلة للتطوير المقدمة من Cassandra.

وتُعد موازنة الأحمال وتوافر التطبيق من الجوانب المهمة لحالة الاســتخدام الخاصة التــي نحن بصددها. وقد أعطت حلول DataStax Enterprise لمخطط مشروع eBay المرونة التي يحتاج إليها لتصميم نظام يُمَكن أي طلب مستخدم من الوصول إلى أي مركز بيانات، مع وجود مركز بيانات يحتوي على مجموعة واحدة من DataStax Enterprise يشــمل هذه المراكز. وتســاعد هذه الســمة في التصميم على موازنة تحميل المستخدم الوارد والقضاء على أي تهديد محتمل للتوقف عن العمل. وتستطيع eBay إجراء تحليل عالي السرعة مع القدرة على الحفاظ على مركز بيانات منفصل يعمل على تشــغيل عقد هــادوب من نفس حلقة DataStax Enterprise (انظر الشــكل ٧-٧) وذلك فضلًا عن خط بيانات الأعمال التي تعمل على تزويد صفحات الويب التي يزورها العملاء.



Source: Data Stax.

شكل ٧-٧: تطوير مركز متعدد البيانات التابع ل: eBay

وتشمل حالة الاستخدام الأخرى شركة Hunch (وهي شركة شقيقة لشركة eBay) "رسم الأذواق" عناصر ومســتخدمي eBay، والذي يقــدم توصيات للعملاء بناءً على اهتمامات المســتخدمين. ويمثل موقع eBay على الويب رســمًا بيانيًا أساسيًا بين جميع المستخدمين والعنــاصر المعروضة للبيع. حيث يتــم التقاط جميع الأحداث (المزايدة، والشراء، والبيع، والسرد) من خلال أنظمة وتخزينها كرســم بياني في Cassandra. ويرى التطبيق أكثر من ٢٠٠ مليون من الكتابات يوميًا ويحمل أكثر من ٤٠ مليار قطعة من البيانات.

وتستخدم eBay أيضًا DataStax Enterprise للعديد من حالات استخدام السلاسل الزمنية والتي تعتبر معالجة كميات ضخمة من البيانات بشكل فوري أمرًا له أولوية قبل كل شيء. وهذا يتضمن تســجيل وتتبع الإشــعار المتنقل (في كل مرة يرســل فيها eBay

إشــعارًا إلى جوال أو أي جهاز فإنه يتم تســجيله في Cassandra)، بالإضافة إلى الكشف عن حالات الغش، وتســجيل طلب / استجابة SOA للتنزيل، وتحليلات وسجلات خادم RedLaser (وهي شركة أخرى شقيقة لشركة eBay).

ومن خلال الاطلاع على جميع حالات الاســتخدام التي ذُكرت، يكون الاشــتراط الشائع هو مدة التشغيل. وتدرك eBay تمامًا حاجتها إلى الحفاظ على أعمالها في مكانة متقدمة بالإضافة إلى الانفتاح من أجل أعمالها، وتلعب DataStax Enterprise دورًا أساســيًا في ذلك من خلال دعمها لمجموعات عالية التوافر. يقول Jay Patel المهندس المخطط لمشروع eBay: «علينا أن نكون على استعداد لتخطي الكوارث طوال الوقت. إنه لشيء رائع حقًا أن يسمح Cassandra بمراكز بيانات متعددة نشطة حيث يمكننا قراءة وكتابة بيانات في أي مكان وفي أي وقت».

أسئلة للمناقشة:

١- لماذا يحتاج eBay إلى حل البيانات الضخمة؟

٢- ما هي التحديات، وما هو الحل المقترح، وما هي النتائج التي تم التوصل إليها؟

Source: DataStax. Customer case studies. datastax.com/resources/casestudies/eBay (accessed July 2016).

حالة عملية ٧-٤

فهم جودة ودقة معلومات دعم الرعاية الصحية على تويتر

يســتطيع جميع مســتخدمي شــبكة الإنترنت اليوم المســاهمة بالمعلومات فضلًا عن الاســتفادة مــن المعلومــات أيضًا. وتُســتخدم هذه القــوة بطرق مختلفة. ويســتطيع المستخدمون على منصات الشبكات الاجتماعية مثل Twitter، نشر معلومات عن أوضاعهم الصحية بالإضافة إلى تلقي المســاعدة لأفضل طريقة لعــلاج هذه الحالات الصحية. وقد اســتفسر العديد من المســتخدمين عن جودة المعلومات المنشورة على منصات الشبكات الاجتماعيــة. وعــلى الرغم من أن القــدرة على نشر المعلومات الصحيــة على تويتر تبدو مفيدة للكثير من المستخدمين الذين يستخدمونها بقصد الحصول على الدعم لمرضهم، إلا أنه دائمًا ما يكون هناك شك حول صحة هذه المعلومات، خاصةً عندما يقوم بنشرها أفراد عاديون. وقد سأل العديد من المستخدمين: «كيف أتأكد وأثق في المعلومات التي يقدمها غير الخبراء حول كيفية التصرف في مسألة حيوية مثل حالتي الصحية؟»

ما هي أنواع المســتخدمين الذين يشــاركون ويناقشــون أي نــوع معلومات؟ وهل يقوم المســتخدمون الذين لديهم عدد كبير من المتابعين بمناقشــة ومشاركة نفس نوع

المعلومات التي يناقشـــها ويشاركها المســـتخدمون الذين لديهم عدد أقل من المتابعين؟ يرجع عدد متابعي المستخدم إلى مدى تأثير المستخدم؛ إذ يتم قياس خصائص المعلومات مـــن حيث الجـــودة والموضوعية من التغريدة المنشـــورة. وقد انـــبرى فريق من علماء البيانات لاستكشـــاف العلاقة بين عدد متابعي المستخدم وخصائص المعلومات التي قام المستخدم بنشرها (Asamoah & Sharda, 2015).

الحل:

تم التنقيب في البيانات من منصة تويتر باستخدام API التابع لـ Twitter. وقد قام علماء البيانات بتكييف نموذج اكتشـــاف المعرفة وإدارة البيانات من أجل إدارة وتحليل هذه المجموعة الضخمة من البيانات. وقد تم تحسين النموذج لإدارة وتحليل البيانات الضخمة المستمدة من منصة الشبكة الاجتماعية وتضمن مراحل اكتساب معرفة المجال، وتطوير منصة البيانات الضخمة الملائمة، والحصول على البيانات وتخزينها، بالإضافة إلى تنظيف البيانات، والتحقق من صحتها، وتحليلها، كما تضمن النتائج والنشر.

التقنية المستخدمة:

تم اســـتخراج التغريدات، وإدارتها، وتحليلها باســـتخدام توزيع Cloudera الخاص بـــ Apache Hadoop. ويحتوي إطار Apache Hadoop على العديد من المشـــاريع الفرعية التي تدعم أنواعًا مختلفة من أنشطة إدارة البيانات، مثل دعم المشروع الفرعي Apache Hive للقراءة والكتابة، وإدارة بيانات التغريدات الضخمة. وقد اســـتُخدمت أدوات لتحليل البيانات مثل Gephi والذي اســـتُخدِم لتحليل الشبكات الاجتماعية وR للنمذجة التنبؤية. وقد قاموا بإجراء تحليلين متوازيين؛ تحليل الشبكة الاجتماعية لفهم الشبكة على المنصة واستخراج النص لفهم محتوى التغريدات التي نشرها المستخدمون.

ماذا وُجِد؟

وكما ذكرنا سابقًا، فقد تم تجميع وتحليل تغريدات من كل المستخدمين المؤثرين وغير المؤثرين. وقد أظهرت النتائج أن جودة وموضوعية المعلومات التي ينشرها المستخدمون المؤثرون كانت أعلى من تلك التي ينشرها المستخدمون غير المؤثرين. كما وجدوا أيضًا أن المستخدمين المؤثرين كانوا مسيطرين على تدفق المعلومات في الشبكة وأن المستخدمين الآخريـــن كانوا أكثر جذبًا لمتابعة رأيهم حول موضوع ما. وقد كان هناك اختلاف واضح بين نوع دعم المعلومات التي يقدمها المســـتخدمون المؤثرون مقابل غيرهم، حيث ناقش المســـتخدمون المؤثرون معلومات أكثر موضوعية فيما يتعلق بإدارة المرض كالتشـــخيص،

والأدوية، والعلاجات الرسمية، في حين قدم المستخدمون غير المؤثرين معلومات أكثر حول الدعم العاطفي والطرق البديلة للتعامل مع مثل هذه الأمراض. وبالتالي، فقد كان هناك اختلاف واضح بين المستخدمين المؤثرين وبين غيرهم من المستخدمين غير المؤثرين.

يصور علماء البيانات من منظور غير الخبراء، إمكانية زيادة توفير الرعاية الصحية من خلال مساعدة المرضى على تحديد واستخدام الموارد القيّمة على الويب من أجل إدارة حالتهم المرضية. وساعد هذا العمل أيضًا على تحديد كيف يمكن لغير الخبراء اكتشاف موقع معلومات الرعاية الصحية التي قد لا تكون بالضرورة مفيدة لإدارة أوضاع المرضى الصحية وترشيحها.

أسئلة للمناقشة:

١- ما هو الشاغل الرئيس لعلماء البيانات فيما يتعلق بالمعلومات الصحية التي يتم نشرها على Twitter؟

٢- كيف يضمن علماء البيانات أن تكون المعلومات المقدمة من غير الخبراء والتي تنشر على وسائل التواصل الاجتماعي، تحتوي بالفعل على معلومات صحية قيّمة؟

٣- هل من المنطقي أن يشارك المستخدمون المؤثرون معلومات أكثر موضوعية بينما يكون تركيز المستخدمين غير المؤثرين على المعلومات الشخصية بشكل أكبر؟ ولماذا؟

Sources: Asamoah, D., & Sharda, R. (2015). Adapting CRISP-DM process for social network analytics: Application to healthcare. In AMCIS 2015 Proceedings. aisel.aisnet.org/amcis 2015/bizAnalytics/GeneralPresentations/33/(accessed July 2016). Sarasohn-Kahn, J. (2008). The wisdom of patients: Health care meets online social media. Oakland, CA: California HealthCare Foundation.

أسئلة للمراجعة على القسم ٧-٤:

١- ما هي الخصائص المشتركة لتقنيات البيانات الضخمة الناشئة؟

٢- ماهو MapReduce؟ وماذا يعمل؟ وكيف يقوم بعمله؟

٣- ماهو هادوب؟ وكيف يعمل؟

٤- ما هي المكونات الرئيسة لـهادوب؟ وما هي الوظائف التي يؤديها؟

٥- ما هو NoSQL؟ وكيف يتناسب مع صورة تحليلات البيانات الضخمة؟

٧-٥ البيانات الضخمة ومستودعات البيانات:

ليس هناك شـك في أن ظهور البيانات الضخمة قد أدى إلى تغيير مستودعات البيانات وسوف يستـمر في تغييرها بشكل كبير. وقد كانت مستودعات بيانات المؤسسات (الفصلين ٢ و٣)، حتى وقـت قريب، هي محـور جميع تقنيات دعم القرار. أما الآن فإن عليهم مشـاركة أضواء الوافد الجديـد، والذي يتمثل في البيانات الضخمة. والسـؤال الذي يطرح نفسـه دائمًا هو هل سـيتم استبدال مستودعات البيانات وتقنياتها الأساسية RDBMS بالبيانات الضخمة وتقنياتها التمكينية مثل هادوب. وهل سنشـهد تحديًا بين مسـتودع البيانات والبيانات الضخمة (أو من وجهة نظر التقنية، بين هادوب وRDBMS)؟ وفي هذا القسـم سـوف نوضح سبب عدم وجود أساس لهذه الأسئلة، وعلى الأقل سوف نبرر أن مثل هذا الاختيار ليس انعكاسًا للواقع في هذا الوقت.

وقد شـهد العقد الماضي أو نحو ذلك، تقدمًا ملحوظًـا في مجال أنظمة دعم القرار المعتمدة على الحاسـب، والتي قد يُنسـب إليها الفضل إلى حد كبير في مستودعات البيانات والتطورات التقنية في كل مـن البرامج والأجهزة للحصول على البيانات وتخزينها وتحليلها. وكلما زاد حجم البيانات، تزداد إمكانيات مسـتودعات البيانات. وقد اشـتمل بعض التقدم الخاص بمسـتودعات البيانات على كل من المعالجة المتوازية على نطاق واسـع (الانتقال من واحد أو عدد قليل إلى العديد من المعالجات المتوازيـة)، وشـبكات منطقة التخزين (حلـول التخزين القابلة للتطوير بسـهولة)، وتخزين الحالة الصلبة، ومعالجة قاعدة البيانات، والمعالجة في الذاكرة، والقواعد العمودية (التوجه بالعمود)، وهذا كلـه مجرد غيض من فيض. وقد سـاعدت هذه التطورات في الحفـاظ على الحجم المتزايد للبيانات لتظـل تحت السـيطرة، بينما تقوم بخدمـة احتياجات التحليلات الخاصة بصانعـي القرار بفاعلية. وما سـاعد على تغير المشـهد في السـنوات الأخيرة هو تنوع وتعقيد البيانات، مما جعل مستودعات البيانـات غير قادرة على مسـايرة التقدم. ومن الجدير بالذكـر أن التنوع والسرعة هي التي أجبرت عالم تقنية المعلومات على تطوير نموذج جديد وليس حجم البيانات، حيث يُطلق على هذا النموذج الآن «البيانات الضخمة». والآن بعد أن أصبح لدينا هذان النموذجان - نموذج مستودعات البيانات ونموذج البيانات الضخمة المتنافسـان ظاهريًا على نفس الوظيفـة - لتحويل البيانات إلى معلومات قابلة للتطبيق - فأي نموذج منهما سوف يفرض نفسه؟ وهل هذا سؤال عادل لطرحه؟ أم أننا نفتقد الصورة الكبيرة؟ ولذلك فإننا نحاول في هذا القسم إلقاء بعض الضوء على هذا السؤال المثير للاهتمام.

وكما هو الحال بالنسـبة للعديد من الابتكارات التقنية السـابقة، فقد أثيرت ضجة حول نموذج البيانـات الضخمة وتقنياته التمكينية مثل هادوب وMapReduce. فغير الممارسـين كغيرهم من الممارسين قد طغت عليهم الآراء المتنوعة. ووفقًا لما ذكره كل من Awadallah وGraham (2012)،

فــإن الزاعمين بأن هــادوب يحل محل قواعد البيانــات العلائقية ويصبح هو مســتودع البيانات الجديد لا يملكون البرهان على مزاعمهم. ومن الســهل معرفــة مصدر هذه الادعاءات لأنه يمكن تشــغيل كل من هادوب وأنظمة مســتودع البيانات بالتوازي، كما يمكن توسيع نطاقها إلى أحجام هائلة من البيانات، بدون مشاركة أي أسلوب بنائي. وأما على المستوى المفاهيمي، فقد يعتقد المرء أنها قابلة للتبادل، في حين أنها ليست كذلك، وأن الاختلافات بين الاثنين تفوق أوجه التشابه. ولكن إذا لم تكن قابلة للتبادل، فكيف نقرر متى نستخدم هادوب ومتى نستخدم مستودع البيانات؟

حالات استخدام Hadoop:

كما ســبق أن وضحنا في هــذا الفصل، لقد كان ظهــور هادوب نتيجة للتطــورات الجديدة في تقنيات الحاســب وشــبكة التخزين؛ حيث يوفر هادوب طبقة من البرامج التي تمتد عبر الشــبكة بأكملها، وتحويلها إلى نظام واحد، وذلك باســتخدام المكونات المادية للحاسب كأساس. وبناءً على ذلك، فإن هناك بعض الاختلافات الرئيسة التي تظهر بوضوح في هذا البناء منها:

- يعتبر هادوب مستودعًا فضلًا عن أنه يعمل على تنقية البيانات الأولية.
- يعتبر هادوب أرشيفًا قويًّا واقتصاديًّا ونشِطًا.

ومن ثم، فإن هادوب يستحوذ على طرفي دورة حياة البيانات واسعة النطاق بدايةً عندما تظهر البيانات الأولية، وأخيرًا عندما تتراجع البيانات، ولكن لاتزال هناك حاجة إليها في بعض الأحيان.

١- هــادوب كمســتودع ومكان تنقية: نظرًا لأن أحجام البيانــات الضخمة تصل من مصادر مثل أجهزة الاستشــعار، والآلات، ووســائل التواصل الاجتماعي وتيارات النقر، فإن الخطوة الأولى تتمثل في التقاط جميع البيانات بطريقة دقيقة وتكلفة عملية. فعندما تكون أحجام البيانات ضخمة، فإن الإســتراتيجية التقليدية للخادم المفرد لا تعمل لفترة طويلة. ومن ثَم فإن تدفق البيانــات في HDFS يمنح مخططي المــشروع المرونة التي هم في أمــس الحاجة إليها، حيث إنهــم لا يمكنهم فقــط التقاط ١٠٠ تيرابايــت في اليوم الواحد، بل يمكنهــم أيضًا ضبط تهيئة هادوب للأعلى أو للأسفل لمواجهة تدفق أو انخفاض البيانات بما يتناسب مع مقدارها، وهذا يتحقق بأقل تكلفة ممكنة لكل جيجابايت نظرًا لاقتصاديات المصادر المفتوحة والاستفادة من المكونات المادية للحاســب. ونظرًا لأن البيانات يتم تخزينها على وحدة التخزين المحلية بدلًا من شــبكات منطقة التخزين، فإن الوصول إلى البيانات في هادوب يكون أسرع بكثير في كثير من الأحيان، كما أنه لا يعيق الشبكة بحركة التيرابايتس من البيانات. وبمجرد التقاط البيانات الخام، فإنه يتم استخدام هادوب لتنقيتها؛ حيث يمكن لـهادوب أن يعمل بمثابة محرك موازي

«محرك ETL على المنشــطات»، من خلال الاســتفادة من تقنيات تحويــل البيانات التجارية أو المكتوبة بخط اليد؛ إذ تتطلب العديد من هذه العمليات الخاصة بتحويل البيانات الأولية، تفكيــك البيانات ذات الصيغة الحرة المعقدة وتحويلها إلى تنســيقات منظمة. وهذا ينطبق بشــكل خاص على تيارات النقر (أو سجلات الويب) وتنســيقات بيانات الاستشعار المعقدة. وبنــاءً عــلى ذلك، فإن المبرمج يحتاج إلى فصل الصالح عن الطالــح (تنقية البيانات من كل ما ليس له أهمية) وتحديد ما هو قَيِّم وسط هذا الصخب.

٢- هادوب كأرشــيف نشــط: في مقابلة أُجريت عام ٢٠٠٣ مع ACM، ادعى Jim Gray أنه من الممكن معالجة الأقراص الصلبة كأشرطة. وعلى الرغم من أن أمر الاستغناء عن أرشفة الشريط المغناطيــسي قد يســتغرق عدة ســنوات، إلا أنه يتــم إعادة توجيه بعض أجــزاء عبء العمل الشريطية بالفعل إلى مجموعات هادوب. ويرجع هذا التحول لســببين أساســيين. أولهما، أنه على الرغم من أن أمر مســتودعات البيانات على الشريــط قد يبدو غير مكلف، إلا أن تكلفته الحقيقية تظهر مع صعوبة الاسترجاع. ولا يقتصر الأمر فقط على الساعات - إن لم تكن الأيام - التي تتطلبها البيانات المخزنة في وضع عدم الاتصال من أجل استرجاعها، بل إن لفائف الشريط نفســها عرضة للتدهور مع مرور الوقت، الأمر الذي يجعل من عملية فقدان البيانات حقيقة واقعة ويجبر الشركات على استيعاب هذه التكاليف. ومما يزيد الأمور سوءًا هو تغير تنسيقات الأشرطــة كل عامين، مــما يتطلب من المنظمات إما إجراء عمليــات ترحيل هائلة للبيانات إلى تنسيق الشريط الأحدث أو المخاطرة بعدم القدرة على استعادة البيانات من الأشرطة القديمة.

وأمــا الأمر الثاني، فقد تبــين أن هناك قيمة لحفظ البيانات التاريخية على الإنترنت مع إمكانية الوصــول إليها بســهولة. وكما هو الحال في مثال تيار النقر، فــإن الاحتفاظ بالبيانات الأولية على قرص (بكرات) اللف لفترة أطول يسهل على الشركات إعادة النظر في البيانات عندما يتغير السياق والقيود الجديدة التي تحتاج إلى تطبيقها. إن عملية البحث عن آلاف الأقراص باستخدام هادوب تكون أسرع وأكثر ســهولة من اللف عبر مئات الأشرطة المغناطيسية. وبالإضافة إلى ما سبق، فإنه مع استمرار تضاعف كثافات الأقراص كل ١٨ شهرًا، يصبح من المُجدي اقتصاديًّا بالنسبة للمنظمات أن تحتفظ بالبيانات الخام أو المنقحة الخاصة بالعديد من الأعوام في HDFS. ومن ثَم، فإن شبكة تخزين هادوب تكون مفيدة في كل من المعالجة الأولية للبيانات الخام ومستودعات البيانات على المدى الطويل. فهو يُعد بحق «أرشــيفًا نشِــطًا» نظرًا لأنه لا يقوم بمستودعات البيانات ويحميها فحسب، بل إنه أيضًا يُمكِّن المستخدمين من استخلاص القيمة بسرعة وسهولة وبشكل دائم.

حالات استخدام مستودعات البيانات:

لقد أصبحت قائمة المميزات المتوفرة في مســتودع البيانات مذهلة بدرجة كبيرة، بعد مرور ما يقرب من ٣٠ عامًا من الاستثمار، والتنقية، والنمو. فقد بُنيت على تقنية قواعد البيانات العلائقية باستخدام المخططات ودمج أدوات ذكاء الأعمال، وتتمثل الاختلافات الرئيسة في هذه البنية في:

- أداء مستودع البيانات.
- دمج البيانات التي توفر قيمة الأعمال.
- أدوات ذكاء الأعمال التفاعلية للمستخدمين النهائيين.

١- أداء مستودع البيانات: إن الفهرسة الأساسية، الموجودة في قواعد البيانات المجانية، مثل MySQL أو Postgres، لهي ســمة نموذجية تُستخدم بغرض تحسين زمن استجابة الاستعلام أو فرض قيود عــلى البيانات؛ حيث تُمكن النماذج الأكثر تقدمًا مثــل: طرق العرض الفعلية، وفهارس الانضمام الإجــمالي، والفهارس المكعبة، وفهارس الانضمام المتفرق، من تحقيق مكاســب متعددة للأداء في مســتودعات البيانات. ومع ذلك، فإن تحسين الأداء الأكثر أهمية حتى الآن المُحَسن القائم على التكلفة؛ حيث يقوم المُحَسن بفحص SQL الواردة ويقوم بدراسة خطط متعددة من أجل تنفيذ كل طلب بحث في أسرع وقت ممكن، حيث يستطيع تحقيق ذلك من خلال مقارنة طلب SQL لتصميم قاعدة البيانات وإحصائيات بيانات شــاملة والتي تســاعد على تحديد أفضل مزيج من خطوات التنفيذ. ويبدو المُحَسن في جوهره وكأنما لديه مبرمج عبقري يقوم بفحص كل استعلام وضبطه من أجل الحصول على أفضل أداء. وبسبب عدم وجود مُحَسن أو إحصائيات ديموغرافية للبيانات، فإن الاســتعلام الذي يمكن تشغيله في دقائق قد يستغرق عدة ساعات، حتى مع وجود العديد من الفهارس. ولهذا السبب، فإن موردي البيانات يقومون باستمرار بإضافة أنواع جديدة من الفهارس، والتقسيم، والإحصائيات، بالإضافة إلى مميزات للمُحَسن. وعلى مدى الثلاثين عامًا الماضية، كان كل إصدار للبرامج عبارة عن إصدار أداء. وكما ســنلاحظ في نهاية القســم الخاص بــهادوب، أنه (أي هادوب) يتفوق على مستودعات البيانات التقليدية في مجال أداء الاستعلام.

٢- دمج البيانات التي توفر قيمة الأعمال: وهي الالتزام بالإجابة على أسئلة الأعمال الأساسية في قلب أي مســتودع. وتمثل البيانات المتكاملة، الأســاس الوحيد المطلوب لتحقيق هذا الهدف. وأما ســبب وجود مستودعات البيانات فيتمثل في سحب البيانات من مجالات تابعة متعددة بالإضافة إلى العديد من التطبيقات إلى مســتودع واحد. ويتعين على مصممي نماذج البيانات ومهندسي اســتخراج، وتحويــل، وتنزيل (ETL) مــع البيانات الوصفيــة (ميتاداتا)، وأدوات تنظيــف البيانات، والعمل الدقيق أن يقوموا بترشــيد تنســيقات البيانــات، وأنظمة المصدر،

والمعنى الدلالي للبيانات لجعلها مفهومة وموثوقة، وهذا يساعد على إنشاء مفردات مشتركة داخل الشركة بحيث يتم قياس وفهم المفاهيم الأساسية مثل: «العميل»، و«نهاية الشهر»، و«المرونة السعرية» بشكل موحد. الجدير بالذكر أن تجميع البيانات وتنظيفها ودمجها كما هي في مستودع البيانات لا يتم إلا في مركز تقنية المعلومات بأكمله، وليس في مكان آخر.

٣- أدوات ذكاء الأعمال التفاعلية: تتيح أدوات ذكاء الأعمال من مثل: MicroStrategy، وTableau، وIBM Cognos، وغيرها من الأدوات للمستخدمين التجاريين إمكانية الوصول المباشر إلى أفكار مستودع البيانات. فبدايةً، يستطيع مستخدم الأعمال إنشاء تقارير بالإضافة إلى إمكانية إنشاء تحليلات معقدة بسرعة وسهولة باستخدام هذه الأدوات. ونتيجة لذلك، فإن هناك اتجاهًا في العديد من مواقع مستودع البيانات نحو الخدمة الذاتية للمستخدم النهائي. كما يستطيع مستخدمو الأعمال بكل سهولة طلب تقارير أكثر مما توفره تقنية المعلومات. غير أن الأهم من الخدمة الذاتية، هو أن يكون المستخدمون على دراية تامة بالبيانات؛ إذ يمكنهم تشغيل تقرير، واكتشاف ما فاتهم من مقياس أو فرز (الفلترة)، بالإضافة إلى إجراء تعديل وتشغيل التقرير الخاص بهم مرة أخرى، كل ذلك في غضون دقائق. ولا شك أن هذه العملية تؤدي إلى حدوث تغييرات كبيرة في فهم المستخدمين للأنشطة التجارية وعملية صنع القرار. وأول هذه التغييرات، هو توقف المستخدمين عن طرح أسئلة تافهة ويبدأون بطرح أسئلة إستراتيجية أكثر تعقيدًا. وبشكل عام، فإنه كلما كان التقرير أكثر تعقيدًا وإستراتيجية، زادت الأرباح والإيرادات ووفورات التكاليف التي يجنيها المستخدم، وهذا من شأنه أن يحول بعض المستخدمين إلى «مستخدمين أقوياء» في الشركة؛ حيث يصبح هؤلاء الأفراد عباقرة في إثارة قيمة الأعمال من البيانات بالإضافة إلى توفير معلومات إستراتيجية قيمة للموظفين التنفيذيين. الجدير بالذكر أن كل مستودع بيانات في أي مكان يحتوي على: من ٢ إلى ٢٠ من المستخدمين الأقوياء.

المجالات الرمادية (أي واحد من الاثنين سوف يقوم بأداء المهمة):

على الرغم من وجود العديد من المجالات التي تميز واحدًا عن الآخر، فإن هناك أيضًا مجالات رمادية لا يمكن فيها تمييز مستودع البيانات عن هادوب بوضوح. ففي هذه المجالات، يمكن للأداة إما أن تكون الحل المناسب وإما أن تقوم بأداء مماثل أو وظيفة غير جيدة على المهمة التي في متناول اليد. ويعتمد اختيار واحدٍ دون الآخر على متطلبات وتفضيلات المنظمة. وفي العديد من الحالات، يعمل كل من هادوب ومستودع البيانات معًا في سلسلة إمداد المعلومات، وكما هو الحال في كثير من الأحيان، تكون إحدى الأداتين هي الأفضل لتحمل عبء عمل معين (Graham & Awadallah, 2012) ويوضح الجدول (٧-١) النظام الأساسي المفضل (واحد مقابل الآخر، أو احتمال المساواة) في إطار عدد من المتطلبات المرصودة الشائعة.

جدول ٧-١: متى يجب استخدام أي منصة Hadoop مقابل DW

المتطلبات	مستودع البيانات	هادوب
وقت استجابة منخفض وتقارير تفاعلية بالإضافة إلى OLAP.	☑	
الالتزام يكون متطلبًا SQL ٢٠٠٣ ANSI.	☑	☑
معالجة أو استكشاف البيانات الخام غير المهيكلة.		☑
المحفوظات على شبكة الإنترنت بديلًا للشريط.		☑
التنظيف عالي الجودة والبيانات المتسقة.	☑	☑
من ١٠٠ إلى ١٠٠٠ من المستخدمين المتزامنين.	☑	☑
اكتشاف علاقات غير معروفة في البيانات.		☑
منطق عملية معقدة متوازية.	☑	☑
تحليل CPU المكثف.	☑	
النظام والمستخدمون وحوكمة البيانات.		☑
العديد من لغات البرمجة المرِنة التي تعمل بالتوازي.		☑
استكشافات sandbox غير المقيدة، وغير الخاضعة للحكم.		☑
تحليل البيانات المؤقتة.	☑	
الأمان الشامل والامتثال التنظيمي.	☑	☑

الجمع بين هادوب ومستودع البيانات:

هنـــاك العديد من الســـيناريوهات المحتملة والتي بموجبها يكـــون الجمع بين كل من هادوب وتقنيات مســـتودعات البيانات القائمة على DBMS العلائقية في الاستخدام، أكثر منطقية. وفيما يلي بعض هذه السيناريوهات (White, 2012):

١- اســـتخدام هادوب لتخزين وأرشفة البيانات متعددة الهياكل: وحينئذٍ يمكن استخدام موصل لـــ DBMS العلائقية من أجـــل التنقيب في البيانات المطلوبة من هادوب لتحليلها بواســـطة

DBMS العلائقيــة. فإذا كانت DBMS العلائقية تدعم وظائــف MapReduce، فإن هذه الوظائف يمكن اســتخدامها للقيام بعملية الاستخراج. وعلى سبيل المثال، فإن محول Aster-Hadoop يَسْــتَخْدِم وظائف SQL-MapReduce لتوفير تنزيل سريع للبيانات ثنائية الاتجاه بــين كل من HDFS وAster Database. وعندئذٍ يمكــن تحليل البيانات التي تم تنزيلها في قاعدة بيانات Aster باستخدام كل من SQL وMapReduce.

٢- استخدام هادوب لتنقية، وتحويل، و / أو دمج البيانات متعددة الهياكل: ويمكن استخدام موصل مثل محول Aster-Hadoop لاستخراج النتائج من معالج هادوب لـ DBMS العلائقية من أجل تحليلها.

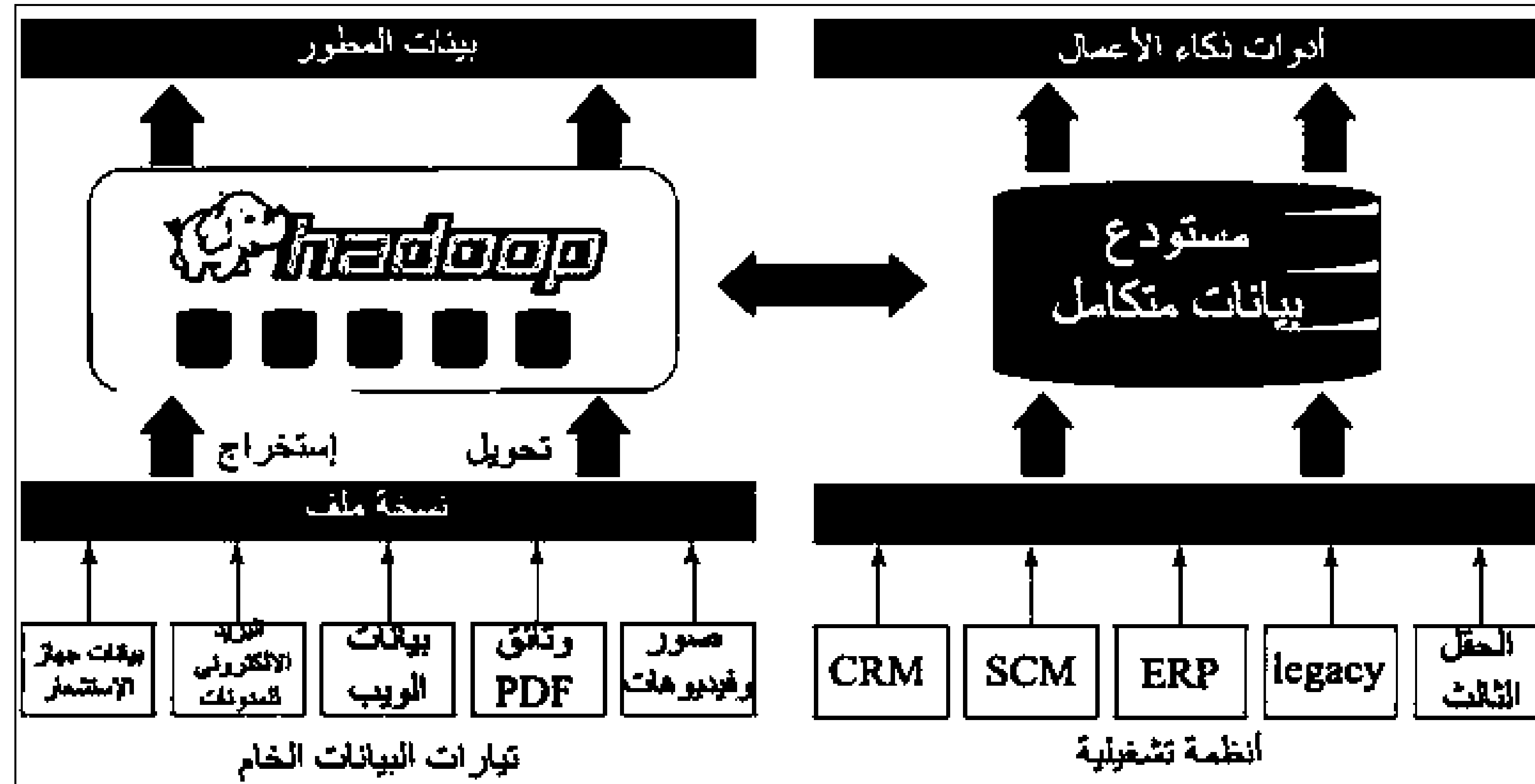
٣- استخدام هادوب لتحليل كميات ضخمة من البيانات متعددة الهياكل ونشر النتائج التحليلية: حيث يعمل هادوب في هذا التطبيق، بمثابة منصة للتحليلات غير أنه يمكن إرجاع النتائج مرة أخرى إلى بيئة مســتودعات البيانات التقليدية، أو مخزن بيانات مشترك لمجموعة العمل، أو واجهة مستخدم مشتركة.

٤- اســتخدام نظــام DBMS العلائقي الــذي يوفر إمكانيــات MapReduce كمنصة حوســبة اســتقصائية: حيث يســتطيع علماء البيانات اســتخدام نظام DBMS العلائقي (مثل: نظام قاعدة بيانات Aster) لتحليل مجموعة من البيانات المهيكلة ومتعددة الهياكل (المحملة من هادوب) باستخدام مزيج يتكون من كل من معالج SQL ووظائف MapReduce التحليلية.

٥- استخدام أداة استعلام Front-end للوصول إلى البيانات وتحليلها: وهنا، يتم تخزين البيانات في كل من هادوب وDBMS العلائقية.

فهــذه الســيناريوهات تدعــم البيئة التي يكون فيهــا كل من هــادوب وDBMSs العلائقية منفصلَــين عن بعضهما البعض كما يتم اســتخدام برنامج الاتصال لتبــادل البيانات بين النظامين (انظر الشــكل ٧-٨). ومــن المرجح أن تتجه الصناعة على مدى الســنوات القليلــة المقبلة نحو تقنيات هادوب المرتبطة بإحكام، بالإضافة إلى تقنيات مســتودع البيانات المستندة إلى نُظُم إدارة قواعد البيانات العلائقية على مســتوى البرامج والأجهزة على حد سواء. حيث يوفر هذا التكامل العديــد من الفوائد، والتي تشــمل عدم الحاجة إلى تثبيت بالإضافــة إلى صيانة أنظمة متعددة، وتقليــل حركة البيانات، وتوفير مخزن بيانات وصفيــة ميتاداتا واحد من أجل تطوير التطبيقات، فضــلًا عــن توفير واجهة واحدة لكل من مســتخدمي الأعــمال والأدوات التحليلية. وقد قدمت المقالة الافتتاحية (القســم ٧-١) مثالًا على كيفية دمج البيانات التي هي من مســتودع البيانات التقليــدي ومجموعتَي البيانات المختلفتين وغير المهيكلتين والمُخزنتين في هادوب من أجل إنشــاء تطبيــق تحليلي للحصــول على إحصاءات حول تفاعلات العميل مع الشركة قبل إلغاء الحســاب.

وبصفتك مديرًا، فإن اهتمامك ينصب على الأفكار التي يمكنك استخلاصها من البيانات، وليس على ما إذا كانت البيانات مخزنة في مستودع بيانات مهيكل أو مجموعات هادوب.



Sources: Teradatacorp.

شكل ٧-٨: التعايش بين Hadoop ومستودعات البيانات

أسئلة للمراجعة على القسم ٧-٥:

١- ما هي التحديات التي تواجه مستودعات البيانات والبيانات الضخمة؟ وهل نشهد نهاية عصر مستودعات البيانات؟ ولماذا نعم أو لماذا لا؟

٢- ما هي حالات الاستخدام للبيانات الضخمة وهادوب؟

٣- ما هي حالات الاستخدام لمستودع البيانات وRDBMS؟

٤- في أي السيناريوهات يمكن الجمع بين هادوب وRDBMS؟

٧-٦ موردو ومنصات البيانات الضخمة:

يتطور المشـــهد الخاص بموردي البيانات الضخمة باعتباره مجالًا تقنيًا جديدًا نســـبيًا. وقد قام عدد مـــن الموردين بتطوير تقســـيمات هادوب الخاصـــة بهم، والتي يعتمد معظمهـــا على توزيع Apache المجاني ولكن بمستويات مختلفة من التخصيص. وهناك اثنان من قادة السوق في مجال التوزيع وهما: Cloudera (cloudera.com) و Hortonworks (hortonworks.com)حيـــث بدأت Cloudera بواســـطة خبراء البيانات الضخمة بما فيهم Doug Cutting منشـــئ هادوب وJeff Hammerbacher عالم البيانات السابق في موقع التواصل الاجتماعي Facebook. أما هورتنورك فقد اشتقت من Yahoo!

حيث تقدم كلتا الشركتين عرض التدريب / الخدمات المدفوعة الأجر على مستوى المؤسسة كما تقدم برمجيات إدارة هادوب مسجلة الملكية، بالإضافة إلى التوزيع الذي تقدمه في الأساس. كما تقوم MapR (mapr.com) - وهي منشأة أخرى في هذا المجال - بتقديم تقسيم هادوب الخاص بها والذي يكمل HDFS بنظام ملفات الشبكة الخاص بها (NFS) من أجل تحسين الأداء. وقد تشاركت شركة EMC Greenplum مع شركة MapR في إصدار تقسيم هادوب الخاص بهما في مايو ٢٠١١. وهذه الأمثلة لا تمثل إلا عددًا قليلًا من الشركات (القديمة والحديثة) التي تزاحم في المشهد التنافسي الخاص بأدوات ومقدمي الخدمات لتقنيات هادوب.

وفي عالم NoSQL، فإن هناك عددًا من الشركات المبتدئة تعمل على تسليم إصدارات معتمدة تجاريًا من الصفات المميزة المختلفة الخاصة بـ NoSQL. فعلى سبيل المثال، تقدم DataStax نسخة تجارية من Cassandra تتضمن دعم المؤسسات والخدمات، فضلًا عن التكامل مع هادوب والبحث عن المؤسسات المجانية عبر Lucene Solr. ويقوم معظم موردو تكامل البيانات الخاصة، بما في ذلك Informatica، وPervasive Software، بالإضافة إلى Syncsort، بعمل جولات في سوق البيانات الضخمة مع روابط هادوب والأدوات المكملة التي تستهدف تسهيل نقل البيانات على المطورين حول وداخل مجموعات هادوب.

كما تشهد طبقة التحليلات في مجموعة البيانات الضخمة أيضًا تطورات مهمة. فعلى سبيل المثال، توجد شركة ناشئة تُسمى Datameer، وتقوم هذه الشركة بتطوير ما تقول عنه بأنه منصة ذكاء الأعمال «الكل في واحد» من أجل هادوب، بينما قام برنامج Tableau المتخصص في تصوير البيانات بإضافة هادوب واتصال مستودع بيانات الجيل التالي بمجموعة منتجاته. وفي الوقت نفسه، فإن شركة EMC Greenplum، لديها كُورَس Chorus، وهو عبارة عن ملعب لعلماء البيانات حيث يمكنهم جمع، وتجربة، وتبادل كميات كبيرة من البيانات من أجل التحليل. وهناك موردون آخرون يركزون على حالات استخدام تحليلية محددة، مثل ClickFox إلى جانب مُحرك تحليلات تجربة عملائها. ويعمل عدد من موردي ذكاء الأعمال التقليديين، وأبرزهم MicroStrategy، على الدمج بين تحليل البيانات الضخمة وكفاءة التقارير في منتجاتهم.

كما أن مساحات تطبيق البيانات الضخمة تنمو أيضًا، حيث تقدم العديد من الشركات تطبيقات مبنية للاستفادة من مجموعات هادوب وإطار MapReduce. وتحتوي الأدوات المجانية مفتوحة المصدر مثل لغة برمجة R على العديد من الدوال المُنفذة للاستفادة من التنفيذ المتوازي من خلال مجموعات. وعلى سبيل المثال، تقدم شركة Treasata تطبيقات البيانات الضخمة كخدمة للعديد من الصناعات.

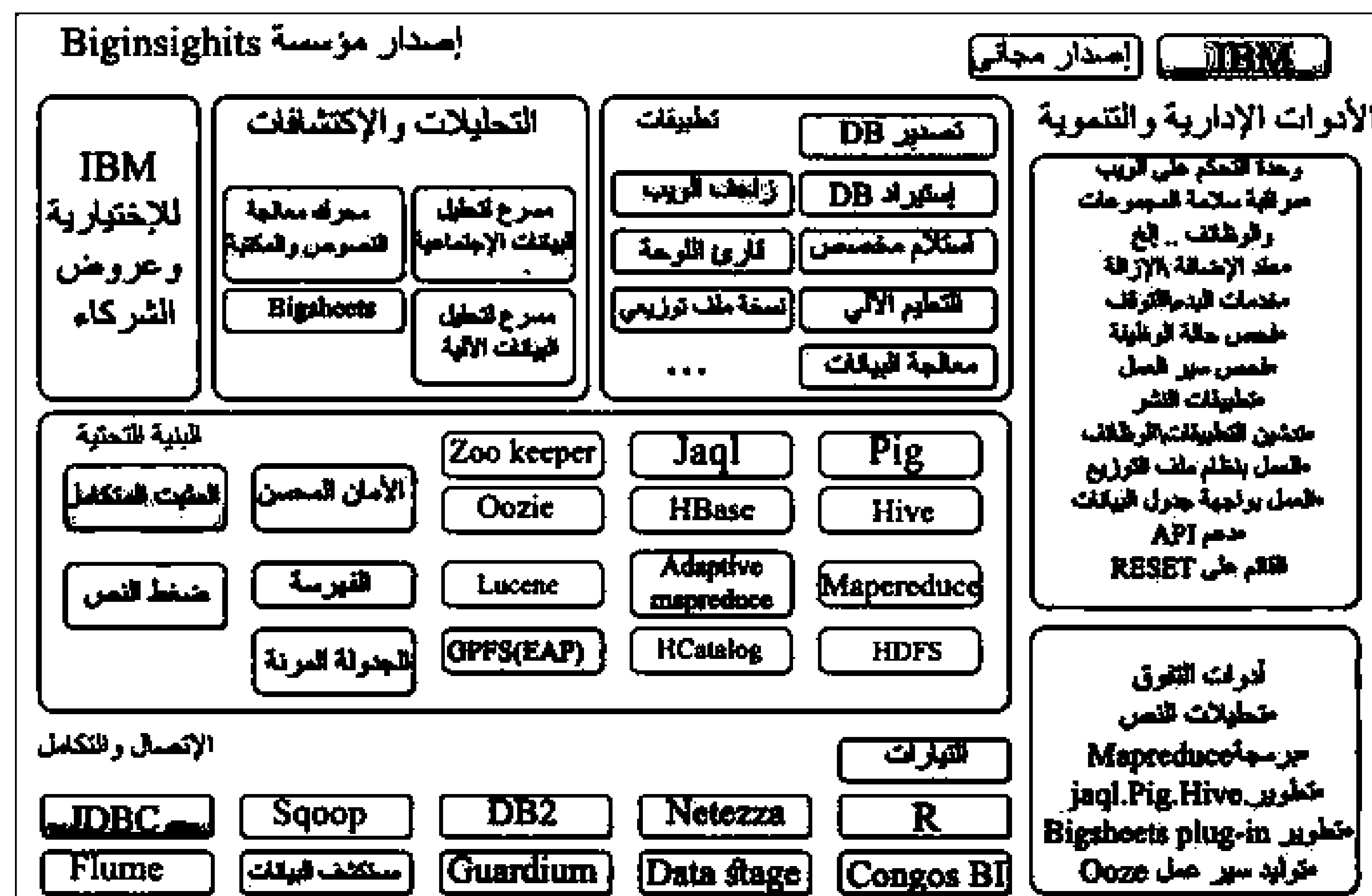
وفي الوقت نفســه، فقد شهد سوق مستودع بيانات الجيل التالي مؤخرًا تجربة دمج كبيرة. فهناك أربعــة من الموردين القادة في هذا المجال وهم Netezza، وGreenplum، وVertica، وAster Data والذيــن تــم اندماجهم مــع IBM، وEMC، وHP، وتيراداتــا، على التوالي. وقد تــم اندماج EMC مــع Dell. كما يلعب المــوردون العمالقة Oracle وIBM أيضًا في مســاحة البيانات الضخمة. فقد تبنَّــت Oracle نهج أدوات البيانات الضخمــة باســتخدام Exadata، وExalogic، وأجهزة البيانات الضخمــة، حيث يقوم جهاز البيانات الضخمة بدمج توزيع Cloudera's Hadoop مع قاعدة بيانات Oracle's NoSQL وأدوات تكامــل البيانات. ويعتمد برنامــج IBM's BigInsights على Apache Hadoop، غير أنه يشتمل على العديد من الوحدات الخاصة والتي تتضمن قاعدة البيانات Netezza، وInfoSphere Warehouse، وأدوات ذكاء الأعمال Cognos، وإمكانيات SPSS للتنقيب في البيانات. كــما يُقدم أيضًا تيــارات IBM InfoSphere، وبرنامج مصمم لتحليل تيارات البيانات الضخمة. ومع نجاح العلامة التجارية Watson analytics، فإن شركة IBM تقوم بطرح العديد من عروض التحليلات الخاصة بها بشــكل عام وعروض البيانات الضخمة بشــكل خاص تحت تصنيف Watson. وقد نتج عن استحواذ تيراداتا على Aster، تقديم منتج مثير للإعجاب في Teradata Aster التي تنفذ عددًا من وظائف التحليلات الشائعة الاستخدام في بيئة البيانات الضخمة. وفيما يلي سنقدم موجزًا عن بيئات كل من IBM's InfoSphere، وTeradata Aster، بالإضافة إلى تقديم حالة عملية قصيرة لكل منهما. وقــد اخترنا تقديم هذين البرنامجين هنا تحديدًا نظرًا لأن كلًا منهما يُعد من البرامج الناجحة تجاريًّا، كما تتوفر لهم مواد تعليمية كثيرة والتي تتضمن البرامج القابلة للتنزيل.

منصة IBM's InfoSphere BigInsights:

مقدمة: يعد InfoSphere BigInsights التابع لشركة IBM منصة قائمة على مشروع Apache Hadoop المجــاني لتحليل البيانــات المهيكلة التقليدية الموجودة في قواعــد البيانات القديمة إلى جانب البيانات شــبه المهيكلة وغير المهيكلة مثل النصوص، والفيديو، والصوت، والصور، ووسائل التواصــل الاجتماعي، وســجلات الويب، وتيارات النقــر. حيث تقوم المنصــة بدمج العديد من تطبيقات MapReduce من خوارزميات التحليلات لتشــغيل تطبيقات متوازية على نطاق واسع وهو مصمم لتوفير تحليلات متقدمة عن تقنية هادوب التي تم تحســينها بشكل خاص من أجل متطلبات تحليل البيانات الضخمة.

أســلوب البنــاء: يعــرض الشــكل (٧-٩) المكونات المختلفــة الخاصة بـــ IBM InfoSphere BigInsights. ويوفــر BigInsights لـــ IBM برنامــج Apache Hadoop القيــاسي، وبالإضافة إلى ذلــك فإنه يوفر أيضًــا تقنيات فريدة ولغات برمجة إلى جانــب التحليلات المدمجة ومُسرعات

التطبيقات (مثل: النص، والجغرافيا المكانية، والسلســلة الزمنية، والتنقيب في البيانات، والتمويل، ووســائل التواصل الاجتماعــي، وحدث الاتصالات، وبيانــات الآلة) لتنفيــذ العمليات المتخصصة بكفاءة من أجل تلبية متطلبات تحليل البيانات الضخمة. فعلى ســبيل المثال، تم تصميم (JAQL JSON Query Language) من أجل تحســين معالجة وتحليل بيانــات ترميز (JSON) وعناصر JavaScript شــبه المهيكلة. وهذه الوظائف مفيدة بشــكل خاص في تحليل تيارات بيانات تويتر. وتمثل BigSheets أداة نمط جدول البيانات التي تدعم استكشاف وتصوير البيانات القابلة للتوسع على جدول SQL الضخم المستقر على HDFS. وتقدم لغة الاستعلام الملحقة (AQL) Annotation Query Language مكتبــات مدمجــة من أجل تحليلات النصوص المتقدمــة عبر كميات هائلة من المســتندات شــبه وغير المهيكلة. أما BigR فهو عبارة عن برنامج للتحليلات واســعة النطاق عــلى هادوب والذي يُمكِّن من الوصول إلى البيانات المســتقرة على HDFS، ومعالجتها، وتحليلها، وتصورها، حسب واجهة المستخدم R. وأخيرًا وليس آخرًا، فإن أدوات InfoSphere BigInsights لـــ Eclipse تُمَكِّــن المطورين من تطوير برامــج لتعمل عــلى InfoSphere BigInsights، حيث تتضمن هذه البرامج معالجات، ومولدات الكود، وبيئة اختبار لتبســيط جهود تطوير التطبيقات. ومع توفر كل هذه الأدوات في برنامج واحد مستمر، يمكن تطوير ونشر التطبيق بسرعة في كتالوج BigInsights القائم على الويب، ومن ثَم استخدام وحدة تحكم خاصة بالويب لنشر التطبيق على مجموعات هادوب والسماح للمستخدمين الذين لديهم الصلاحية أن يصلوا إليه.



شكل ٧-٩: بنية IBM Infosphere

كيف تبدأ: يمكن تنزيل تجارب BigInsights من الرابط:

http://www.ibm.com/analytics/us/en/technology/hadoop/hadoop-trials.html

كما أن إصدار البدء السريع من BigInsights متاح للتنزيل مجانًا عبر

http://www.ibm.com/developerworks/downloads/im/Biginsightsquick/

وهــو يعالــج مجموعة أحادية أو متعددة من المصدر المفتــوح Hadoop. كما يتطلب نظامًا يحتوي على ذاكرة وصول عشوائي تبلغ سعتها 16 GB على الأقل، بالإضافة إلى معالج رباعي النواة، و50 GB مساحة حرة.

وأفضل مكان لبدء تعلم ما يخص InfoSphere BigInsights هو من خلال مركز IBM للمعلومات:

http://www.ibm.com/support/knowledgecenter/SSPT3X_2.1.2/com.ibm.swg.im.InfoSphere.Biginsights.tut.doc/doc/tut_Introduction.html.

وتقــدم الحالة العمليــة (٧-٥) مثالًا يوضح كيــف يمكننا دمج البيانات مــن مصادر متعددة لتحليل تقارير الانفلونزا.

حالة عملية ٧-٥

استخدام وسائل التواصل الاجتماعي للتنبؤ الفوري بنشاط الانفلونزا

تفــرض الأمــراض المُعدية عبئًا كبيرًا على نظــام الصحة العامــة في الولايات المتحدة الأمريكيــة. وقــد أدى ظهور فيروس نقــص المناعة البشرية / الإيــدز HIV/AIDS في أواخر الســبعينيات، وفــيروس الانفلونــزا H1N1 في عام ٢٠٠٩، ووبــاء H3N2 خلال موســم شــتاء ٢٠١٢ - ٢٠١٣، وتفشّي فيروس إيبولا Ebola في عــام ٢٠١٥، بالإضافة إلى فــيروس زيكا Zika الــذي انتشر في عــام ٢٠١٦، إلى إثبات قابلية النــاس للإصابة بمثل هــذه الأمــراض المُعدية. ومن الناحيــة الفعلية، فإنه يحدث تفــشّي للانفلونزا كل عام بأشــكال مختلفة وينتج عنها آثار متباينة. ويبلغ متوسط التأثير السنوي لتفشّي فيروس الانفلونزا الموســمية في الولايات المتحدة ٦١٠٦٦٠ شخصًا فَقَد الحياة من أعمار مختلفة غير محصورة، و٣,١ مليون يوم في المستشــفى، و٣١,٤ مليون زيارة للعيادات الخارجية، وإجــمالي ٨٧,١ مليار دولار كعِبءٍ اقتصادي. ونتيجة لهذا الاتجاه المتزايد، فقد ظهرت على الســاحة في السنوات الأخيرة أســاليب جديدة لتحليل البيانات بالإضافة إلى ظهور تقنيات قادرة على اكتشــاف مثل هذه الأمراض وتتبعها ورســم خرائطها وإدارتها. وقد أعطت أنظمة المراقبة الرقمية بوجه خاص، وعودًا بقدرتها على اكتشــاف أنماط البحث

عن الصحة العامة وتحويل هذه الاكتشافات إلى إستراتيجيات قابلة للتطبيق.

وقد بين هذا المشروع أنه يمكن اســتخدام وسائل التواصل الاجتماعي كطريقة فعالة للكشــف المبكر عن تفشي الانفلونزا. فقد استخدمنا منصة البيانات الضخمة لاستخدام بيانات Twitter من أجل مراقبة نشــاط الانفلونــزا في الولايات المتحدة. وقد تضمنت أســاليبنا في تحليل البيانات الضخمة اســتخراج كل من الزمان، والمــكان، والنص. فقد قمنــا في التحليل الزمني بفحص ما إذا كان يمكن بالفعل تكييف بيانات Twitter للتنبؤ الفــوري بمدى تفشّي الانفلونزا. كما قمنا في التحليــل المكاني، بتخطيط تفشي الانفلونزا للخاصيــة الجغرافية المكانية من بيانات تويــتر لتحديد النقاط التي تؤثر فيها الانفلونزا بشــكل فعال. وقد تم إجراء تحليلات النص للتعرف على الأعراض والعلاجات الشــائعة للانفلونزا والتي تم ذكرها في التغريدات.

وقد تم اســتخدام منصــة IBM InfoSphere BigInsights لتحليل مجموعتين من بيانات نشاط الانفلونزا؛ حيث استُخدمت بيانات تويتر لمراقبة تفشّي الانفلونزا في الولايات المتحدة، بالإضافة إلى اســتخدام مستودع بيانات Cerner HealthFacts لتتبع اللقاءات السريريــة في العالم الحقيقي. وقد انتقل حجم ضخم من التغريدات المتعلقة بالانفلونزا مــن تويــتر باســتخدام Twitter Streaming API حتــى تم اســتيعابه في مجموعات هــادوب. وبمجــرد تلقي البيانات بنجاح، اســتُخدمت أداة JSON Query Language (JAQL) لمعالجــة وتحليل بيانات ترميز عناصر JavaScript الشــبه منظمة (JSON). وبعد ذلك، تم استخدام Hive لهيكلة البيانات النصية وفصل المعلومات من أجل تحليل الموقع المكاني والزماني والتصوير في R. كما تم تنفيذ عملية التنقيب في البيانات بالكامل باســتخدام وظائف MapReduce. وقد استخدمنا الحزمة BigR لتقديم نصوص R عبر البيانــات المخزنة في HDFS. حيث مكنتنا الحزمة BigR من الاســتفادة من الحســاب المتــوازي لنظام HDFS كما مكنتنا من إجراء عمليات MapReduce. وقد اســتخدمت مكتبات API لخرائط جوجل كأداة رسم خرائط أساسية لتصوير مواقع التغريد.

وقــد أظهــرت النتائج التــي توصلنا إليهــا أن التكامل بــين كل من وســائل التواصل الاجتماعي والسجلات الطبية يستطيع أن يكون مكملًا ذا قيمة لنظم المراقبة القائمة. كما أكدت نتائجنا أن الحركة المرتبطة بالانفلونزا على وسائل التواصل الاجتماعي ترتبط ارتباطًا وثيقًا بتفشّي الانفلونزا الفعلية. وقد أظهر ذلك أيضًا باحثون آخرون (St Louis, Zorlu, 2012; Broniatowski, Paul, & Dredze, 2013). وقد أجرينــا تحليل سلســلة زمنية

للحصول على الارتباط المتبادل المكاني والزماني بين الاتجاهين (٩١٪) ولاحظنا أن مواجهات الانفلونزا الإكلينيكية تأتي خلف المشـاركات عبر الإنترنت. وبالإضافة إلى ذلك، فقد كشف تحليل موقعنا عن عدة مواقع عامة نشـأت من خلالها أغلب التغريدات. وهذه النتائج يمكنها مسـاعدة المسـؤولين عن الصحة والحكومات على تطوير نمـاذج تنبُّؤ أكثر دقة في الوقت المناسب أثناء حدوث التفشِّي، وإبلاغ الأفراد حول المواقع التي يجب تجنبها خلال تلك الفترة الزمنية.

أسئلة للمناقشة:

١- لماذا تسـتطيع وسائل التواصل الاجتماعي أن تكون بمثابة مؤشر مبكر على تفشِّي الانفلونزا؟

٢- ما هي المتغيرات الأخرى التي قد تساعد في التنبؤ بمثل هذه التفشيات؟

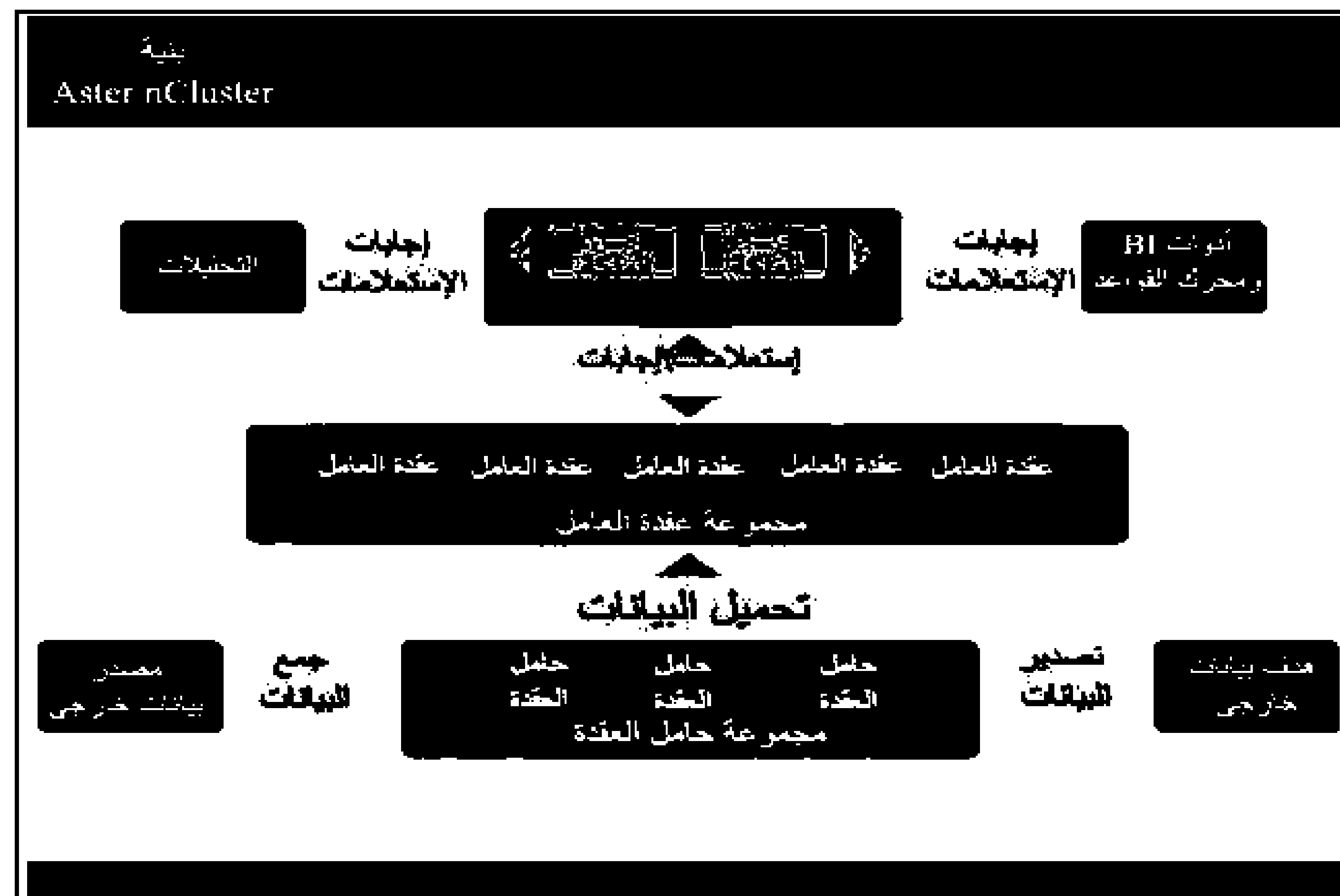
٣- لماذا تُعد هذه المشكلة مثالًا جيدًا يمكن حله باستخدام تقنيات البيانات الضخمة المذكورة في هذا الفصل؟

Sources: Zadeh, A. H., Zolbanin, H. M., Sharda, R., & Delen, D. (2015). Social media for nowcasting the flu activity: Spatial- temporal and text analysis. Business Analytics Congress, Pre-ICIS Conference, Fort Worth, TX. Broniatowski, D. A., Paul, M. J., & Dredze, M. (2013). National and local influenza surveillance through Twitter: An analysis of the 20122013- influenza epidemic. PloS One, 8(12), e83672. Moran, P. A. (1950). Notes on continuous stochastic phenomena. Biometrika, 1723-.

منصة Teradata Aster:

مقدمـة: إن Teradata Aster هي منصة بيانات ضخمة للتخزين الموزع ومعالجة مجموعات البيانات الكبيرة متعددة الهياكل. وقد استُخدِمت هذه المنصة لتحسين التسويق، وكشف الاحتيال، كما استُخدِمت في التحليلات الرياضية، وتحليل الشبكات الاجتماعية، بالإضافة إلى تحليلات البيانات الآليـة، وتحليلات الطاقة، وتحليلات الرعاية الصحية، والعديد من التطبيقات الأخرى. وقد قامت Teradata Aster بموازاة العديد من وظائف التحليلات التقليدية والمتقدمة، كما أن لديها قدرات لإجراء تحليل السلاسـل الزمنية، والتحليل الإحصـائي، وتحليل الكتلة، والتنقيب في النص، بالإضافة إلى التنقيب في قاعدة الترابط، وتحليل الشبكات الاجتماعية، والتحليلات المرئية، وتحليلات الموقع، والتحليـلات التنبؤية؛ حيث تقوم بإجراء كل ذلك بطريقـة موزعة. وبالإضافة إلى حزم التحليلات التقليدية، فإن Teradata Aster لديها أيضًا العديد من حزم التحليلات الجديدة والفريدة لتحليل المسار، كما أنها متوافقة أيضًا مع لغات البرمجة الأخرى مثل R، وPython، وJava.

أسـلوب البناء: لقد تم تكييف Teradata Aster مع أسـلوب بناء master-slave الخاص بـ Apache Hadoop. وهو يتكون من عقدة ملكة وعقد عامل متعددة وهي تعادل كلًا من عقد الاسم وعقد البيانات في هادوب، على التوالي. ويقدم الشكل (٧-١٠) أسلوب بناء TD Aster.



Sources: Teradata Corp.

شكل ٧-١٠: بنية Teradata Aster

وتقوم عقدة الملكة في الطبقة العليا بإدارة النظام، والمخطط، ومعالجة الأخطاء، وتوزيع الحساب على العاملين. فهي تقوم بتنسـيق الاسـتعلامات وإرجاع نتائج الاسـتعلام. أما الطبقة الوسطى من البنـاء فهي تحتوي على العمال، حيث يقوم العمال بمسـتودعات البيانات باسـتخدام عامل التكرار الذي يعينه المسؤول، ومن ثمَّ، فإنهم يتحملون الأخطاء. ويتفاعل العمال مع بعضهم البعض لمعالجة الاستعلامات التي تطلبها الملكة. ويتكون المستوى الثالث من البناء من عقدة مسؤول التنزيل والتي تستقبل البيانات من أطراف ثالثة مثل نظم المعلومات.

وتشتمل بنية TD Aster على متجر ملف أستر Aster File Store والذي يمكنه استيعاب بيانات متعددة مثل سـجلات الويب، وبيانات جهاز الاستشـعار، وبيانات سـجل الآلة. وهو متوافق مع HDFS وأنظمـة الملفات التقليدية الأخرى. ويمكن أيضًا توصيـل المجموعة بقواعد بيانات أخرى مثل Oracle DB، وTeradata Warehouse، بالإضافة إلى Hive من خلال الروابط الأخرى المتاحة.

وبالنسـبة لتحليل البيانات، فإن المسـتخدم يمكنه كتابة الاسـتعلامات التقليدية لكل من SQL، و SQL-MapReduce (SQL-MR)، وSQL-Graph (SQL-GR). ويُعـد SQL-MR إطـارًا لـ TD Aster والذي يُمكن أداء MapReduce من وظائف التحليلات في قاعدة بيانات Aster. وبالمثل، فإن

SQL-GR يعتبر إطار عمل من أجل تمكين معالجة محرك الرسم البياني في قاعدة بيانات Aster. ويتم تنفيذ عملية الاستعلامات المكتوبة في SQL أو SQL-SQL-MR أو SQL-GR تلقائيًّا بشكل متوازي عبر المجموعة. وللوصول إلى قواعد البيانات وكتابة الاستعلامات، فإنه يمكن للمستخدمين إما استخدام أداة Aster Command (ACT)أو استخدام تطبيق من جانب العميل يسمى Teradata Studio.

وتشتمل الحزمة الكاملة من TD Aster على أداة تستند إلى متصفح الويب للتصوير المرئي تسمى AppCenter. حيث يمكن استخدام AppCenter لإنشاء أنواع مختلفة من التصويرات المرئية مثل مخطط Sankey، ومخطط sigma، ومخطط chord، وهيكل الشجرة، ومخطط المجموعة الهرمية، وسحابة الكلمة، والمخطط الشريطي، والمخطط الدائري، وغيرها من الرسوم البيانية الإحصائية التقليدية الأخرى.

كيف تبدأ: يتوفر الإصدار السريع من Teradata Aster للتنزيل مجانًا. حيث يمكن تنزيل TD Aster Express من هذا الرابط (https://aster-community.teradata.com). وهو يتكون من جهازين افتراضيين: ملكة واحدة وعامل واحد. كما أنه يتطلب جهازًا لا يقل عن ٤ GB من الرامات «ذاكرة الوصول العشوائية». ويمكن للمستخدم تنزيل ١٧ GB من البيانات في Aster Express. كما يمكن الوصول إلى دليل البدء من خلال بوابة شبكة جامعة تيراداتا (www.teradatauniversitynetwork.com). وللوصول إلى الوثائق والتعليمات والتمارين ومجموعات البيانات من أجل تعلم Teradata Aster فإنه يجب على المرء إنشاء حساب على موقع TUN على الويب.

وقد سبق أن رأينا أمثلة على استخدام Teradata Aster في المقالة الافتتاحية في الفصل الأول في تحليلات الرياضة وأيضًا في بداية هذا الفصل. وتوفر الحالة العملية(٧-٦) مثالًا آخر على استخدام Teradata Aster.

حالة عملية ٧-٦

تحليل أنماط المرض من مستودع بيانات السجلات الطبية الإلكترونية

لقد تم منح مركز الإبداع في الأنظمة الصحية في جامعة أوكلاهوما الحكومية، مستودعًا ضخمًا للبيانات من قِبل شركة Cerner Corporation، والتي هي عبارة عن مزود السجلات الطبية الإلكترونية الرئيسة (EMRs)، للمساعدة في تطوير التطبيقات التحليلية. حيث يشتمل مستودع البيانات على EMRs في زيارات أكثر من ٥٠ مليون من المرضى الفريدين من نوعهم عبر مستشفيات الولايات المتحدة (١٩٩٥ - ٢٠١٤). كما أنه يشمل أكثر من ٨٤ مليون زيارة قاهرة وطارئة وإسعافية. فهي أكبر قاعدة بيانات علائقية وحيدة في هذه الصناعة والتي تشتمل على سجلات شاملة مع الصيدليات،

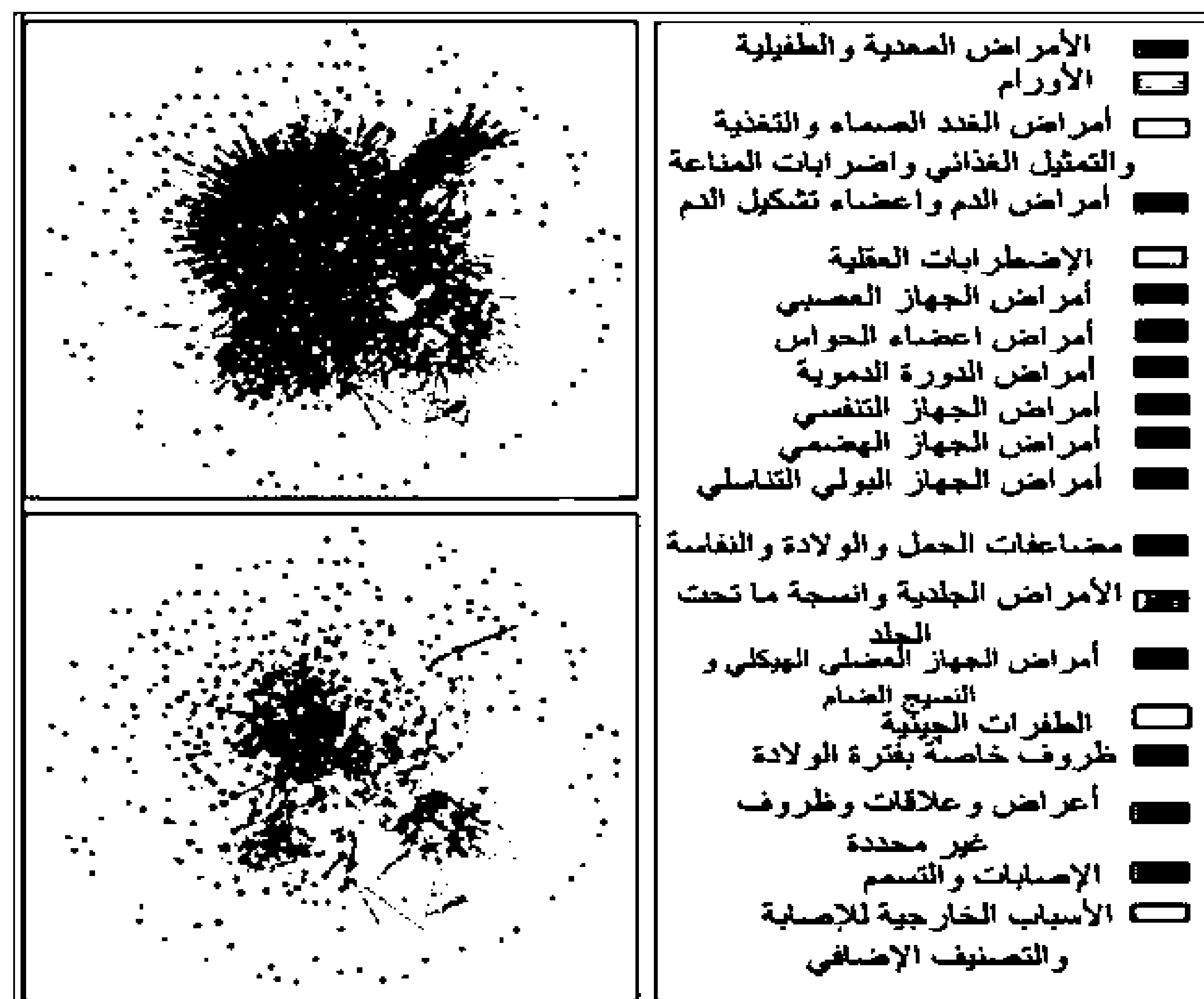
والمختبرات، والأحداث السريرية، والقبول وبيانات الفوترة. كما تشتمل قاعدة البيانات علـى أكثر من ٢,٤ مليار نتيجة معملية وأكثر من ٢٩٥ مليون طلب لما يقرب من ٤٥٠٠ علاج بالاسـم والعلامة التجارية. حيث تُعد هـذه القاعدة واحدة من أكبر مجموعات البيانات من نوعها غير محددة الهوية على أرض الواقع والمتوافقة مع HIPAA.

ويمكن اسـتخدام EMRs لتطوير تطبيقات تحليلية متعددة. وأحد هذه التطبيقات هو فهم العلاقات بين الأمراض على أسـاس المعلومات حول الأمراض المتزامنة (التي تأتي في وقـت واحد) المتقدمة في المـرضى. فعندما يصاب مريض واحـد بعدة أمراض، فإن هذه الحالة تسـمى اعتلال مشـترك. ومن الممكن أن تكون الأمراض المصاحبة مختلفة عبر المجموعات السـكانية. وفي هذه الحالة العملية، قامت مجموعة بحثية في جامعة أوكلاهوما الحكومية بعمل مقارنة بين الأمراض المصاحبة في المرضى من المناطق الحضرية والأمراض المصاحبة في المرضى من المناطق الريفية.

وللمقارنـة بين الأمـراض المصاحبة، فقد تـم تطبيق نهج تحليل الشـبكة. وتتألف الشـبكة من مجموعة محددة من العناصر تسـمى العُقد، وترتبط هذه العُقد ببعضها البعض من خلال الحواف؛ حيث تمثل الحافة علاقة محددة بين العُقد. ومن أكثر الأمثلة شيوعًا على الشبكة، مثال شبكة صداقة يتواصل فيها الأفراد مع بعضهم البعض في حالة ما إذا كانوا أصدقاء. وعلى نحو مماثل، فهناك شـبكات أخرى مشـتركة مثل شـبكات الحاسـب، وشبكات صفحات الويب، بالإضافة إلى شـبكات الطرق، وشبكات المطارات. ولمقارنـة الأمراض المصاحبة، فقد تم تطوير الشـبكات الخاصة بالأمراض في المرضى من المستشـفيات الريفية والحضرية. وقد اسـتُخدِمت المعلومات المتعلقة بالأمراض والتي طورها كل مريض خلال زيارات المستشـفى لإنشـاء شـبكة للمرض. وقد بلغ إجمالي عدد زائري المستشـفيات في المستشفيات الحضرية ٦٦ مليون، في حين بلغ إجمالي عدد الزائرين في المستشفيات الريفية ١ مليون شخص. ولإدارة مثل هذه المجموعة الضخمة من البيانات، تم اسـتخدام منصة البيانات الضخمة Teradata Aster. كما استُخدمت أطر SQL، وSQL-MR، وSQL-GR والمدعومة بواسـطة Aster. أيضًا فقد اسـتُخدم Aster AppCenter، وGephi لتصوير الشبكات مرئيًا.

ويعرض الشـكل (٧-١١) شبكات الإصابة المرضية المشتركة الريفية والحضرية. حيث تمثـل العُقد في هذه الشـبكات، الأمـراض المختلفة المصنفة بحسـب التصنيف الدولي للأمراض، المراجعة التاسـعة، التعديل السريري (ICD-9 CM)، مجمعة على مسـتوى ثلاثة أرقام، حيث يتم الربط بين اثنين من الأمراض في حالة ما إذا كانا مرتبطين بشـكل

كبير أو بشكل مرضي ($p < 0.01$). وكلما زاد حجم العقدة، دل ذلك على زيادة الإصابة بالمرض. الجدير بالذكر أن كثافة شبكات الإصابة المرضية الحضرية أكثر من كثافة الشبكة الريفية؛ حيث يبلغ عدد العقد والحواف في الشبكة الحضرية ١,٠٤٣ و٢٢,٠٢٩ على التوالي، في حين أن عدد العقد والحواف في الشبكة الريفية يبلغ ٩٩٣ و٢,٠٧٣ على التوالي، مما يدل على أن تشخيص المرضى في المستشفيات الريفية غالبًا ما يشير إلى وجود عدد أقل من الأمراض المتزامنة. وتعرض التصويرات فرقًا واضحًا بين نمط الأمراض المتقدمة في المناطق الحضرية والريفية، مما يستدعي العديد من الأسئلة المتعلقة بالسياسات الطبية والاجتماعية والتي تتطلب المزيد من البحث والتحليل. وعلى الجانب الآخر، فقد قمنا بتضمين هذا التحليل لزيادة الوعي أيضًا بقضية أخرى. وقد لاحظنا مبكرًا أن مجموعة البيانات المستخدمة لهذا التحليل شملت حوالي ٦٦ مليون لقاء مع المرضى في المناطق الحضرية و١ مليون لقاء فقط مع المرضى في المناطق الريفية. وعلى الأرجح فإن هذا الاختلاف الكبير يرجع لعدم قدرة معظم المستشفيات الريفية على تحمل تكاليف نظام السجلات الطبية الإلكترونية الكبرى مثل Cerner، وبالتالي فإن البيانات تميل نحو المستشفيات الحضرية، غير أن أي أفكار نتجت عن هذه العينة سوف يتم التشكيك فيها. وكما ذكرنا في الفصل الرابع، فإنه ربما تم سحب عينة متناسبة من المناطق الحضرية وتمت مقارنتها بسجلات المرضى الريفيين.



شكل ٧-١١: شبكات الأمراض المشتركة بين الحضر والريف

وســتتحمل نظم قواعد البيانات التقليدية جهــدًا في تجهيزها بكفاءة للبيانات الضخمة. وقد قام Teradata Aster بتحليل البيانات التي تحتوي على معلومات عن ٨٤ مليون زيارة و٢٠٠ مليون ســجل سريع وســهل إلى حد ما. وغالبًا ما يتم اقتراح تحليل الشــبكة على اعتبار أنــه طريقة واحدة لتحليل مجموعات البيانات الضخمة؛ فهو يساعد على فهم البيانات في صورة واحدة. في هذه الحالة العملية، تبين شبكة الإصابة المرضية المشتركة العلاقة بين الأمراض في مكان واحد.

أسئلة للمناقشة:

١- لمــاذا يمكن أن تكون الإصابة المشــتركة للأمراض مختلفة بين المستشــفيات في الريف وفي الحضر؟

٢- ما هي مســألة الفرق الكبير بين مواجهات المــرضى في كل من المناطق الريفية والحضرية؟

٣- ما هي المكونات الرئيسة للشبكة؟

٤- أين أيضًا يمكنك تطبيق نهج الشبكة؟

Source: Kalgotra, P., & Sharda, R. (2016). Rural versus urban comorbidity networks. Working Paper, Center for Health Systems and Innovation, Oklahoma State University.

كما يزداد الدور الذي تلعبه السحابة في سوق البيانات الضخمة. ويقوم كل من Amazon وGoogle بدعم عمليات نشر هادوب في عروض السحابة العامة. أما Amazon Elastic MapReduce وGoogle Compute Engine على التوالي، فإنهما يقومان بتمكين المســتخدمين من توســيع النطاق وتقليل حجم المجموعات بســهولة حســب الحاجة. وتقوم مايكروسوفت (Microsoft) بدعم توزيع هادوب الخاص بـهورتنورك في سحابة Azure. وفي الفصل التالي سوف نناقش العروض القائمة على السحابة.

كما ينتهج أيضًا موردون آخرون نهج البيانات الضخمة من زاوية التحليلات المرئية. وكما يشير Magic Quadrant أحدث إصدارات Gartner، فإن هناك نموًا كبيرًا في ذكاء الأعمال والتحليلات في الاستكشــافات والتحليلات المرئية. وتقوم الشركات الكبيرة مثل SAS، وSAP، وIBM، بجانب شركات صغــيرة غير أنها مســتقرة مثــل Tableau، وTIBCO، وQlikView، بتقديم حالة قوية لتحليــلات الأداء العــالي الموجودة في منصات تصوير المعلومات. وتقــدم رؤية فنية (٧-٣) عددًا قليلًا من العوامل المســاعدة الرئيســة لتحقيق النجاح مع البيانــات الضخمة والتحليلات المرئية. وربمــا يكــون SAS هو الدافع لها بقوة أكثر من أي شيء آخر مع منصته التي ظهرت مؤخرًا وهي SAS للتحليلات المرئية، وقد ســبق تقديمه في الفصل الثاني. وباســتخدام العديد من التحسينات

الحسابية يستطيع برنامج التحليلات المرئية SAS تحويل عشرات الملايين من سجلات البيانات إلى رســومات بيانية معلوماتية خلال بضع ثوانٍ فقط باســتخدام المعالجة المتوازية الواسعة (MPP) والحوسبة داخل الذاكرة.

رؤية فنية ٧-٣

كيف تنجح مع البيانات الضخمة

من البيت الأبيض وحتى منزلك يصعب العثور على منظمة أو شخص لديه اليوم بيانات أقل من عام مضى، حيث تتزايد خيارات قاعدة البيانات، كما يتطور ذكاء الأعمال نحو عهد جديد من التحليلات على مستوى المنظمة، بالإضافة إلى أن كل شيء يتحرك ولا يبقى واقفًا في مكانه. فالمنظمات التي استطاعت تكييف هيكل البيانات الخاص بها وعملياتها بنجاح مــن أجل معالجة الخصائص الثلاثة للبيانات الضخمة - الحجم، والتنوع، والسرعة - تعمل على تحسين الكفاءة التشغيلية، وتزايد الإيرادات، بالإضافة إلى تمكين نماذج أعمال جديدة. والحقيقة أن معدل التغيير سوف يتحقق فقط مع كل المنظمات التي تهتم بوضعها الذي يتعلــق بالابتــكار في كل ما يخــص البيانات. إذًا ما الذي يجب عــلى الشركات فعله حتى تحقق النجاح مع البيانات الضخمة؟ وفيما يلي بعض من التوصيات الصناعية:

١- التبســيط: إنــه لمن الصعب أن تتــم متابعة جميع موردي قواعــد البيانات الجدد، وجميع المشــاريع المجانية، وجميع مقدمي خدمات البيانات الضخمة، بالإضافة إلى أن كل هذه الأشياء سوف تزداد ازدحامًا وتعقيدًا في الأعوام المقبلة. ومن ثم، فهناك حاجة ماســة للتبسيط. فمن الضروري اتباع نهج إستراتيجي من خلال توسيع نطاق أنظمــة معالجة المعاملات العلائقيــة والاتصال عبر الإنترنت الخاصة بك إلى واحد أو أكثر من البرامج الجديدة المُســتضَافة داخل المؤسســة، أو خيــارات قاعدة البيانات القائمــة على الخدمة والتي تعكس احتياجات صناعتــك ومنظمتك على أكمل وجه، ومن ثم اختيار منصة ذكاء الأعمال الذي يدعم الاتصالات المباشرة للعديد من قواعد البيانــات وتنســيقات الملفات في الوقت المناســب. ومن الجديــر بالذكر أن اختيار أفضل مزيج من بدائل الحلول لكل مشروع (ما بين الاتصال المباشر بقواعد البيانات السريعة وتلقي مستخرجات البيانات في داخل محرك التحليلات في الذاكرة لتعويض أداء قواعــد البيانــات البطيئة أو المثقلــة بالأعباء) هو أمر بالــغ الأهمية لنجاح أي مشــاريع بيانات ضخمة. فعلى سبيل المثال: تتألف بنية تحليلات البيانات الضخمة من تيراداتا (وهي إحدى أشــهر شركات مستودعات البيانات)، وهادوب (وهو الحل

الأكــثر نجاحًا لتحدي البيانات الضخمة)، بالإضافــة إلى Tableau (وهو أحد موردي حلــول التحليلات البصرية غزيرة الإنتاج). ويمكن لموظفي eBay تصوير الأفكار من أكثر من ٥٢ بيتابايتس من البيانات. ويستخدم eBay حل التحليلات المرئية بواسطة Tableau لتحليل مدى ملاءمة البحث وجودة موقع eBay، الذي يقوم بمراقبة أحدث ملاحظــات العملاء وقياس ميولهم، بالإضافة إلى تحقيق التقارير التشــغيلية لأنظمة مستودعات البيانات، وقد ساعد كل ذلك على نجاح الثقافة التحليلية داخل eBay.

٢- التعايش: إن اســتخدام نقاط القوة لكل منصة قاعدة بيانات، وتمكينها من التعايش في بنية بيانات مؤسستك لهو أمر ضروري. وهناك العديد من الكتابات التي تحدثت عــن ضرورة الحفاظ عــلى التعايش بين مســتودعات البيانــات التقليدية وقدرات المنصات الجديدة وتعهده بالرعاية.

٣- التصوير: بالنســبة للشركات الرائدة في أبحاث التحليلات مثل شركتَي Forrester، و Gartner، فإن المؤسســات تتحرى أن يكون لديها منصات متقدمة لتصوير البيانات من أجل أن تكون أدوات أساســية تُمكِّنهم من مراقبة الأعمال، والعثور على الأنماط، واتخــاذ إجراءات لتجنب التهديدات واقتناص الفرص. وتســاعد التحليلات البصرية المنظمات في الكشــف عن الاتجاهات، والعلاقات، والعيوب عن طريق التحول المرئي من خلال كميات كبيرة جــدًا من البيانات. ولتجربة التحليل المرئي خصائص معينة؛ فهي تتيح لك القيام بشيئين في أي لحظة، وهما:

- تغيير البيانات التي تبحث عنها بشــكل فــوري. وتكمن أهمية هذا الأمر في كون الأسئلة المختلفة تتطلب بيانات مختلفة.
- تغيير الطريقة التي تنظر بها إلى البيانات بشــكل فوري. ويُعد ذلك مهمًا لأن كل عرض قد يجيب عن أسئلة مختلفة.

وهــذا المزيج يُحدث التجربة الاســتطلاعية المطلوبة لأي شــخص للإجابة على الأســئلة بسرعــة. وهذه التجربــة في جوهرها تؤدي إلى أن يصبح التصويــر البصري امتدادًا طبيعيًا لعملية التفكير التجريبي.

٤- التمكين: تســير البيانــات الضخمة جنبًا إلى جنب مع خدمــة ذكاء الأعمال الذاتية، ووفقًــا لمجموعــة Aberdeen التي نُشِرت مؤخرًا، فإن «قيمــة التحليلات والبيانات الضخمــة تزداد إلى أقصى حد». ومن الجدير بالذكر أن المنظمات التي لديها بيانات ضخمــة تمتلك فرصــة تفوق نســبة ٧٠٪ في احتمالية أن يكون لديها مشــاريع /BI BA أكثر من المنظمات الأخــرى والتي يتم توجيهها من قِبل مجتمع الأعمال، وليس

مــن قِبل مجموعة تقنية المعلومات. وعبر مجموعة من الاســتخدامات من معالجة مشــاكل الأعمال الجديدة، وتطوير منتجات وخدمات جديدة بالكامل، والعثور على الذكاء القابل للتطبيق في أقل من ســاعة، ومــزج البيانات من مصادر متباينة تجعل البيانات الضخمة تطلق العنان لما هو ممكن من خلال تطبيق التحليلات.

٥- التكامل: يُعد تكامل ومزج البيانات من مصادر مختلفة جزءًا أساســيًا من تحليلات البيانــات الضخمــة لمنظمتك. فالمنظمات التــي يمكنها مزج مصــادر بيانات أولية، علائقيــة مختلفة، وشــبه مهيكلة، في الوقت المناســب، وبــدون تكاليف دمج أولية باهظة الثمن، هي المنظمات ذات أفضل قيمة من البيانات الضخمة. وبمجرد حدوث الدمــج والمزج، فإن هيكل البيانات (مثل: جداول البيانــات، أو قاعدة البيانات، أو مســتودع البيانات، أو نظام ملفات مجاني مثل هادوب، أو جميعها في نفس الوقت) يصبح غير مهم؛ وهذا يعني أنك لست بحاجة إلى معرفة تفاصيل كيفية مستودعات البيانات لطرح الأســئلة والإجابة عليها. وكما رأينــا في حالة عملية (٧-٤)، أن حملة أوباما وجدت طريقة لدمج وســائل التواصل الاجتماعــي، والتقنية، وقواعد بيانات البريد الإلكتروني، وجمع البيانات، وبيانات السوق الاستهلاكية لعمل ميزة تنافسية.

٦- الحوكمــة: لطالما كانت حوكمة البيانات مشــكلة صعبة في مجــال تقنية المعلومات، ومــع ظهور البيانات الضخمة أصبح الأمر أكثر حــيرة. وهناك أكثر من ٨٠ دولة لديها قوانين لخصوصية البيانات. وقد حدد الاتحاد الأوروبي سبعة «مبادئ خصوصية للملاذ الآمن» لحماية بيانات مواطنيهم الشــخصية. ففي سنغافورة بدأ تفعيل قانون حماية البيانات الشــخصية في يناير ٢٠١٣. وفي الولايات المتحدة الأمريكية، يؤثر Sarbanes-Oxley عــلى جميع الشركات المدرجة في البورصة، ويضع قانون HIPAA (قانون نقل التأمين الصحي والمُســاءلة وقابلية النقل) المعايير الوطنيــة في مجال الرعاية الصحية. ويختلــف التوازن الصحيــح بين التحكم والتجريب تبعًا لاختــلاف التنظيم والصناعة. ويبدو أن أفضل ممارسات إدارة البيانات الرئيسة تساعد في إدارة عملية الحوكمة.

٧- التبشــير: فمــن خلال الحصول عــلى دعم من واحد أو أكثر مــن الرعاة التنفيذيين، يستطيع المُبشِرون مثلك الحصول على الكرة المتداولة وغرس مجموعة فاضلة: فكلما زاد عــدد الإدارات في منظمتك والتي تحقــق فوائد قابلة للتطبيق، تصبح التحليلات أكثر انتشــارًا عبر منظمتك. وتُعد التحليلات المرئية السريعة والســهلة في الاستخدام هي المفتاح الذي يفتح الباب أمام تبنِّي التحليلات والتعاون على نطاق المنظمة.

Sources: Lampitt, A. (2012). Big data visualization: A Big deal for eBay. infoworld.com/d/Big-data/Big-data-visualization-Big- deal-ebay-208589 (accessed August

2016). Tableau white paper. (2012). 7 Tips to Succeed with Big Data in 2013. cdnlarge.tableausoftware.com/sites/default/files/whitepapers/7-tips-to-succeed-with-Big-data-in-2013.pdf (accessed August 2016).

أسئلة للمراجعة على القسم ٧-٦:

١- ما المقصود بمشهد موردي البيانات الضخمة؟ ومن هم اللاعبون الكبار؟

٢- في رأيك كيف سيتغير مشهد موردي البيانات الضخمة في المستقبل القريب؟ ولماذا؟

٣- ما دور التحليلات المرئية في عالم البيانات الضخمة؟

٧-٧ البيانات الضخمة وتحليلات التيار:

تُعد السرعة من أهم الخصائص الرئيسة التي تميز البيانات الضخمة إلى جانب خاصيتَي الحجم والتنوع، اللتان رأيناهما سابقًا في هذا الفصل، وتشير السرعة هنا إلى السرعة التي يتم بها إنشاء البيانات وتدفقها إلى بيئة التحليلات. وتبحث المنظمات عن وسائل جديدة لمعالجة تيار البيانات عندما تأتي للاستجابة بسرعة وبدقة على المشاكل والفرص من أجل إرضاء عملائهم والحصول على ميزة تنافسية. وفي الحالات التي تأتي فيها تيارات البيانات بشكل سريع وبصورة مستمرة، فإنه غالبًا ما تصل طرق التحليل التقليدية التي تعمل مع البيانات المتراكمة سابقًا (أي البيانات في حالة الراحة) إلى قرارات خاطئة بسبب استخدام الكثير من البيانات خارج السياق، أو تصل هذه الطرق إلى القرارات الصحيحة، ولكن بعد فوات الوقت الذي تحتاج المنظمة إلى استخدامها فيه. لذلك؛ فإنه من المهم في العديد من مواقف العمل أن يتم تحليل البيانات بعد وقت قصير من إنشائها و / أو بمجرد أن يتم بثها في نظام التحليلات.

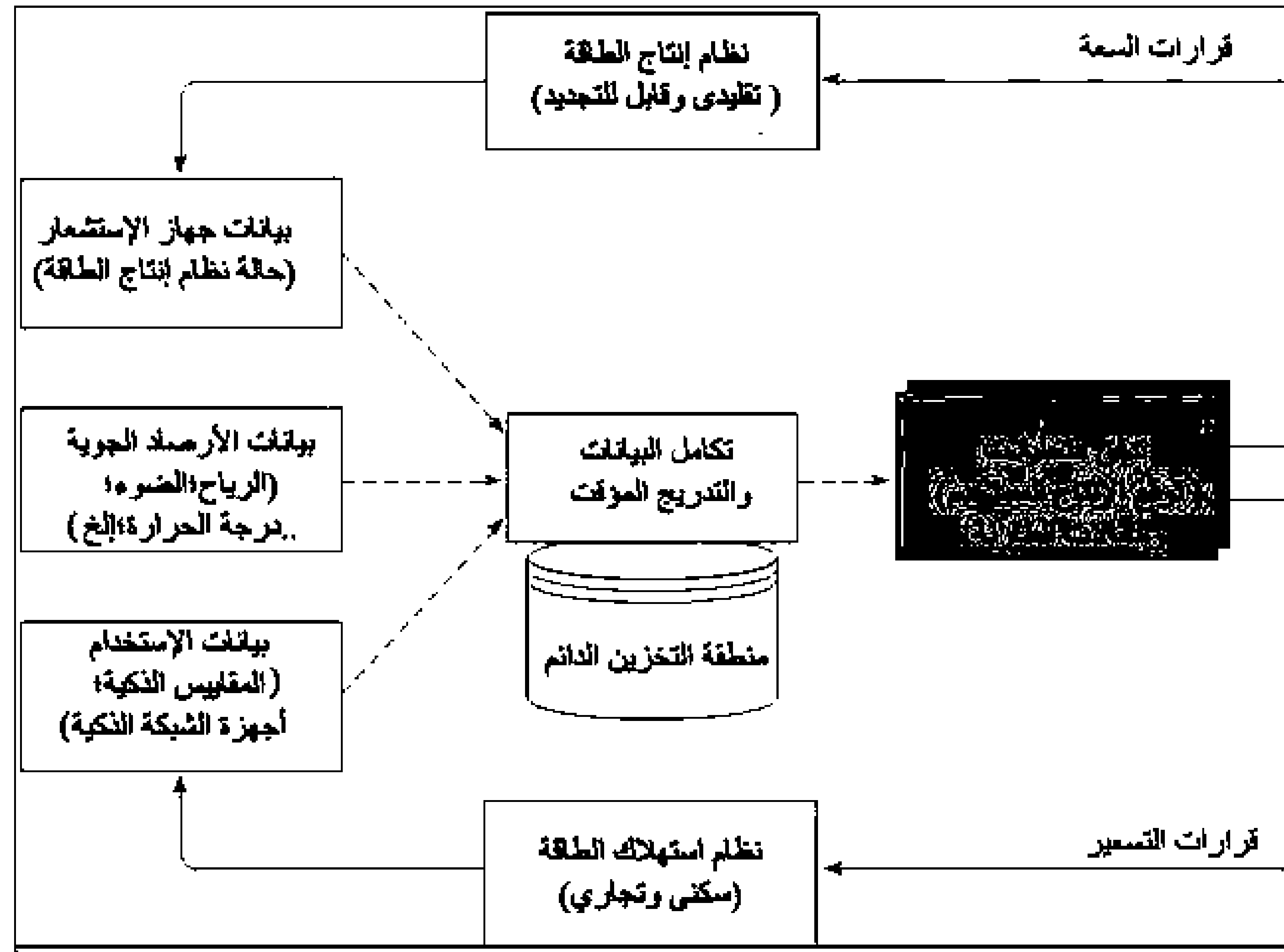
ومن المفترض أن أغلب الأعمال في العصر الحديث تعيش حاليًا على مبدأ أنه من المهم والضروري تسجيل كل جزء من البيانات؛ لأنه ربما يحتوي على معلومات قيمة قد يحتاج إليها العمل الآن أو في وقت ما في المستقبل القريب. وكيفما يزداد عدد مصادر البيانات، فإن نهج «تخزين كل شيء» يصبح أصعب وأصعب، بل إنه قد يكون غير ممكن في بعض الحالات. وفي الواقع، فإنه على الرغم من التقدم التقني، إلا أن سعة التخزين الإجمالية الحالية تختلف كثيرًا عن المعلومات الرقمية التي يتم إنتاجها في العالم. وعلاوةً على ذلك، فإن الكشف المستمر عن التغييرات ذات الدلالة في البيانات بالإضافة إلى الاختلافات المُعقدة للأنماط ضمن نافذة زمنية قصيرة معينة في الوقت المناسب في بيئة الأعمال المتغيرة، يُعد أمرًا ضروريًا للتوصل إلى الإجراءات التي تتلاءم بشكل أفضل مع البيئة الجديدة. ومما سبق يتبين أن كل هذه الحقائق التي ذُكِرت

تُعـد هي المحفز الرئيس لظهور نموذج يُسـمى بتحليلات التيار. وقد ظهر نموذج تحليلات التيار كاسـتجابة لكل هذه التحديات، وبالتحديد، التدفقات غير المحـدودة من البيانات التي لا يمكن تخزينها بشكل دائم ليتم تحليلها لاحقًا، في الوقت المناسب وبكفاءة، بالإضافة إلى تحدي التغيرات النمطية المعقدة التي تحتاج إلى الكشف عنها والتصرف فور حدوثها.

ويُسـتخدم مصطلح تحليلات التيار (ويسـمى أيضًا تحليلات البيانـات في الحركة والتحليلات الفوريـة للبيانات) عادةً من أجل العملية التحليلية الخاصة باسـتخراج معلومات قابلة للتطبيق من بيانات التيار / البيانات المتدفقة بشكل متواصل. ويُعرف التيار بأنه تسلسل مستمر لعناصر البيانـات (Zikopoulos وآخـرون، ٢٠١٣). وغالبًـا ما تُسـمى عناصر البيانـات في مجموعة ما بالصفوف tuples، بمعنى قاعدة البيانات العلائقية، حيث يشـبه tuple صفًا من البيانات (سِجل، موضوع، مثال). ومع ذلك، فإنه في سـياق البيانات شـبه المهيكلة أو غير المهيكلة، فإن tuple هو عبارة عن تجريد يمثل حزمة من البيانات، والتي يمكن وصفها بأنها مجموعة من السمات لموضوع معـين. فإذا كان tuple بحد ذاته غير وافٍ بالمعلومـات الكافية للتحليل أو الارتباط أو العلاقات الجماعية الأخرى بين الصفوف المطلوبة، فعندئذٍ يتم اسـتخدام نافذة للبيانات تتضمن مجموعة مـن الصفـوف tuples. هذه النافذة من البيانات عبارة عن عـدد / تتابع محدود من الصفوف tuples، حيـث يتم تحديث النوافذ باسـتمرار كلما توافرت بيانـات جديدة. ويتم تحديد حجم النافذة بناءً على النظام الجاري تحليله. وقد أصبح انتشار تحليلات التيار متزايدًا لسببين. أولهما، أن قيمة وقت العمل أصبحت في تناقص مسـتمر، وأما ثانيهما، أننا لدينا الوسائل التقنية لالتقاط ومعالجة البيانات أثناء إنشائها.

وقد تم تطوير بعض التطبيقات التي تُعد أكثر تطبيقات تحليلات التيار تأثيرًا في صناعة الطاقة، وتحديدًا من أجل أنظمة الشـبكات الذكية (سلسـلة إمدادات الطاقة الكهربائية). فالشـبكات الذكية الجديدة لا يمكنها فقط إنشـاء ومعالجة تيارات متعددة للبيانات في الوقت المناسـب من أجل التوزيع الأمثل للطاقة لمقابلة احتياجات العملاء الحقيقية، بل إنها أيضًا يمكنها إنشاء تنبؤات دقيقة على المدى القصير تهدف إلى تغطية الحاجات المُلحة غير المتوقعة وتوليد الطاقة المتجددة. ويوضح الشـكل (٧-١٢) صورة حالة اسـتخدام عامة لتحليلات التيـار في صناعة الطاقة (تطبيق الشـبكة الذكية النموذجـي)، والغرض من ذلك هو التنبؤ بدقة بالطلـب على الكهرباء وإنتاجها في الوقت المناسـب باستخدام بيانات التيار التي توفرها العدادات الذكية وأجهزة استشعار نظام الإنتاج ونماذج الرصد. ويمكن اسـتخدام كل من القدرة على التنبؤ باتجاهات الاستهلاك / الإنتاج في المسـتقبل القريب بالإضافة إلى اكتشاف الحالات الشـاذة في الوقت المناسب من أجل تحسين

قــرارات التوريد (كمية الإنتاج، وموارد الإنتاج المســتخدمة، وتعديل قــدرات الإنتاج على النحو الأمثل) بالإضافة إلى ضبط العدادات الذكية لتنظيم الاستهلاك والتسعير الملائم للطاقة.



شكل ٧-١٢: حالة استخدام تحليلات التيار في صناعة الطاقة

تحليلات التيار مقابل التحليلات الدائمة:

قد يبدو لمعظم الناس أن كلا المصطلَحين التيار والدائم هما الشيء نفســه، كما أنهما يســتخدمان في كثير من الحالات بشكل مترادف، غير أنه يوجد اختلاف بينهما في سياق الأنظمة الذكية (Jonas, 2007). فتحليــلات التيار تنطوي على تطبيق منطق مســتوى المعاملات إلى المشــاهدات الفورية؛ حيث إن القواعــد المطبقــة على هذه الملاحظات تأخــذ في اعتبارها الملاحظات الســابقة طالما أنها وقعت في النافذة المقررة؛ وتحتوي هذه النوافذ على حجم عشوائي (مثل: آخر ٥ ثوانٍ، وآخر ١٠,٠٠٠ ملاحظة). وعلى الجانب الآخر، فإن التحليلات الدائمة تقوم بتقييم كل مشاهدة واردة في مقابل كل المشاهدات الســابقة، حيث لا يوجد حجم نافذة. ومن الجدير بالذكر أن التعرف على كيفية ارتباط المشــاهدة الجديدة بجميع المشاهدات السابقة يُمكِّن من تكوين رؤية في الوقت المناسب.

ولكل مــن تحليلات التيار والتحليلات الدائمة مزاياها وســلبياتها وأماكنها الخاصة في عالم تحليلات الأعمال. فعلى سبيل المثال، أحيانًا تكون أحجام المعاملات كبيرة للغاية ويكون وقت اتخاذ القرار قصيراً جــدًا، مما يؤدي إلى تفضيل عدم الثبات وأحجــام النوافذ الصغيرة، والتي تُترجم إلى تحليلات التيار، في حين أنه عندما تكون المهمة خطيرة ويمكن إدارة وحدات التخزين في الوقت المناســب، فعندئذٍ تكون

التحليــلات الدائمــة هي الخيار الأفضل. وبهذه الطريقة، فإنه يمكن الإجابة على أســئلة من نوع: «ما علاقة ما تعلمته للتو بما كنت أعرفه؟»، «هل هذا شيء مهم؟»، «من الذي يحتاج إلى المعرفة؟».

معالجة الأحداث الحرجة:

إن معالجة الأحداث الحرجة هي طريقة التقاط، وتتبُّع، وتحليل تيارات البيانات للكشف عن الأحداث (التي تكون خارج الأحداث العادية) لأنواع معينة تستحق الجهد المبذول. وتُعد معالجة الأحــداث المعقدة تطبيقًــا لتحليلات التيار والتــي يتم فيها جمع البيانات مــن مصادر متعددة لاستنتاج الأحداث أو أنماط الاهتمام إما قبل حدوثها بالفعل أو بمجرد حدوثها. والهدف من ذلك هو اتخاذ إجراءات سريعة من أجل منع حدوث الآثار الســلبية لهذه الأحداث أو التخفيف منها (مثل: الاحتيال أو اختراق الشــبكات)، أو في حالة وجود بصيص أمل من الفرصة، فيكون الهدف هو الاســتفادة الكاملة من الموقف في الوقت المســموح به (بناءً على سلوك المستخدم على موقع التجارة الإلكترونية، وإنشاء عروض ترويجية والتي يُرجح أن يستجيب لها المُستخدم).

وقــد تقع هذه الأحداث الحرجــة عبر طبقات مختلفة من المنظمة مثــل العملاء المحتملين، أو الطلبات، أو مكالمات خدمة العملاء، أو على نطاق أوسع، والذي قد يكون عناصر إخبارية، أو رسائل نصية، أو منشــورات وســائل التواصل الاجتماعي، أو تغذيات سوق الأوراق المالية، أو تقارير حركة المــرور، أو الأحــوال الجوية، أو غيرها من أنواع الحالات الشــاذة التي قد يكــون لها تأثير كبير على تحسين حالة المنظمة. كما يمكن أيضًا تعريف الحدث بشكل عام على أنه «تغيير الحالة»، والذي قد يتم اكتشافه على أنه قياس يتجاوز بداية محددة في وقت سابق، أو درجة الحرارة، أو قيمة أخرى. وعــلى الرغم من أنــه لا يوجد إنكار لاقتراح القيمة الخاصة بمعالجة الأحداث الحرجة، إلا أنه يتعين انتقاء ما يجب قياسه، ومتى يتم القياس، وكم مرة يتم القياس. وبسبب الكم الهائل من المعلومات المتوفرة حول الأحداث، والتي يشــار إليها أحيانًا باســم سحابة الحدث، فإن هناك احتمال للمبالغة فيها، وفي هذه الحالة فإنها قد تضر بالفعالية التشغيلية للمنظمة، بدلًا من مساعدتها.

التنقيب في تيار البيانات:

يمكن تعريف التنقيب في تيار البيانات، باعتباره تقنية تمكينية لتحليلات التيار، على أنه عملية اســتخراج الأنماط الجديدة وهياكل المعرفة من سجلات البيانات السريعة والمتواصلة. وكما رأينا في فصل التنقيــب في البيانات (الفصل الرابع)، فإن الطرق التقليدية للتنقيب في البيانات تتطلب جمع البيانات وتنظيمها في شــكل ملف مناسب، ثم يتم معالجتها بطريقة متكررة لمعرفة الأنماط الأساسية. وعلى النقيض من ذلك، فإن تيار البيانات هو تدفق مستمر لتسلسل مرتب من النماذج

التي يمكن قراءتها / معالجتها في كثير من التطبيقات الخاصة بالتنقيب في تيار البيانات مرة واحدة فقط أو عدد قليل من المرات باستخدام قدرات محدودة للحوسبة والتخزين. وتشتمل أمثلة تيار البيانات على بيانات جهاز الاستشـعار، وزيارات شبكة الحاسب، والمحادثات الهاتفية، ومعاملات ATM، وبحث الويب، والبيانات المالية. ويعتبر التنقيب في تيار البيانات حقلًا فرعيًا من التنقيب في البيانات، وتعلم الآلة، واكتشاف المعرفة.

ويتمثل الهدف في العديد من تطبيقات التنقيب في تيار البيانات، في التنبؤ بالدرجة أو القيمة الخاصـة بالنـماذج الجديدة في تيار البيانات التي توفر بعض المعرفـة حول عضوية الفئة أو قيم النماذج السـابقة في تيار البيانات. ويمكن اسـتخدام أسـاليب تعلم الآلة المتخصصة (ومعظمها مشـتقة من أسـاليب تعلم الآلة التقليدية) لتعلم مهمة التنبؤ من الأمثلة المصنفة بطريقة آلية. وقد قام كل من Delen، وKletke، وKim (2005)، بتطوير مثال على طريقة التنبؤ، حيث قاموا تدريجيًا ببناء وصقل نموذج لشجرة القرار باستخدام مجموعة فرعية من البيانات في وقت واحد.

أسئلة للمراجعة على ٧-٧:

١- ما هو التيار (في عالم البيانات الضخمة)؟

٢- ما هي الدوافع لتحليلات التيار؟

٣- ما هي تحليلات التيار؟ وكيف تختلف عن التحليلات العادية؟

٤- ما هي معالجة الأحداث الحرجة؟ وكيف ترتبط بتحليلات التيار؟

٥- عرِّف التنقيب في تيار البيانات. ما هي التحديات الإضافية التي يطرحها التنقيب في تيار البيانات؟

٧-٨ تطبيقات تحليلات التيار:

إن اسـتخدام تحليلات التيار يتخذ اتجاهًا متزايدًا بشكل كبير، وذلك نظرًا لقدرته على إحداث رؤية حالية، مما يسـاعد صانعي القرار على التواجد على قمة الأحداث عندما يتوسـعون، الأمر الـذي يتيح للمنظـمات معالجة القضايا قبل أن تتحول إلى مشـكلات. وفيما يلي بعض مجالات التطبيق التي استفادت بالفعل من تحليلات التيار.

التجارة الإلكترونية:

تحـاول العديد من الشركات تحقيق أقصى اسـتفادة من البيانـات التي تجمعها أثناء وجود العميـل على موقع الويب الخاص بها، مثل شركة Amazon، وشركة eBay (من بين العديد من

الشركات الأخرى). حيث يتم تسجيل وتحليل كل صفحة تتم زيارتها، وكل منتج تمت مشاهدته، وكل عملية بحث أُجريت، بالإضافة إلى كل نقرة يتم عملها، وذلك لزيادة القيمة المكتسبة من زيارة المستخدم. فإذا تم القيام بذلك بسرعة، فإن تحليل مثل هذا التيار من البيانات يمكنه تحويل المتصفحين إلى مشترين والمشترين إلى مدمني تسوق. فعندما نقوم بزيارة موقع للتجارة الإلكترونية على الويب، حتى وإن لم نكن أعضاءً فيه، فإننا بعد عدة نقرات، نبدأ في الحصول على عروض أسعار منتجات وحزم مثيرة للاهتمام بشكل كبير. ويتم خلف الكواليس بشكل فوري إجراء التحليلات المتقدمة للبيانات الناتجة عن نقراتنا، وآلاف النقرات من الآخرين، «لفهم» ما نحن مهتمون به (في بعض الحالات، حتى ولو لم نكن نعرف ذلك) وتحقيق أقصى استفادة من هذه المعلومات من خلال تقديم عروض إبداعية.

الاتصالات الهاتفية:

إن حجم البيانات التي تأتي من سجلات تفاصيل المكالمات (CDR) لشركات الاتصالات، أمر مدهش. وعلى الرغم من أن هذه المعلومات قد استُخدمت لأغراض إعداد الفواتير منذ فترة ليست بالقصيرة، إلا أن هناك ثروة من المعرفة المدفونة في عمق هذه البيانات الضخمة والتي تدركها شركات الاتصالات الآن. فعلى سبيل المثال، يمكن تحليل بيانات CDR لمنع تحول العملاء إلى شركة أخرى عن طريق تحديد الشبكات الخاصة بالمتصلين، والمؤثرين، والقادة، والمتابعين داخل هذه الشبكات والعمل على هذه المعلومات بشكل استباقي. وكما نعلم جميعًا، فإن للقادة وللمستخدمين المؤثرين تأثيرًا كبيرًا في تغيير نظرة المتابعين داخل شبكتهم نحو مقدم الخدمة، سواء بشكل إيجابي أو سلبي. وتقوم شركات الاتصالات بتحديد القادة وأصحاب النفوذ والمشاركين في شبكاتهم لإدارة قاعدة عملائهم بشكل أفضل وذلك باستخدام أساليب تحليل الشبكات الاجتماعية. كما يمكن بالإضافة إلى ذلك استخدام هذه المعلومات أيضًا لاستقطاب أعضاء جدد ورفع قيمة الأعضاء الحاليين إلى أقصى حد.

ويمكن دمج التيارات المستمرة من البيانات التي تأتي من CDR، مع بيانات وسائل التواصل الاجتماعي (تحليل المشاعر) من أجل تقييم فعالية الحملات التسويقية. كما يمكن استخدام الأفكار المكتسبة من هذه التيارات الخاصة بالبيانات للتفاعل السريع مع التأثيرات الضارة (التي قد تؤدي إلى فقدان العملاء) أو تعزيز وقع التأثيرات الإيجابية المشاهدة في هذه الحملات (والتي قد تؤدي إلى تعظيم عمليات الشراء للعملاء الحاليين واستقطاب عملاء جدد). وعلاوةً على ذلك، فإنه يمكن تكرار عملية اكتساب الأفكار من CDR من أجل شبكات البيانات التي تستخدم سجلات تفصيلية لبروتوكول الإنترنت. ونظرًا لقيام معظم شركات الاتصالات بتوفير كلا النوعين من

هذه الخدمات، فإن التحسين الشامل لجميع العروض والحملات التسويقية قد يؤدي إلى مكاسب استثنائية في السوق. وتُعد حالة عملية (٧-٧) مثالًا على كيفية فهم Salesforce.com لعملائه بشكل أفضل من خلال الاعتماد على تحليل تيارات النقر.

حالة عملية ٧-٧

استخدام Salesforce لتيار البيانات لتحسين قيمة العميل

وسعت شركة Salesforce خدمات السحابة التسويقية الخاصة بها لتشمل النتائج التنبؤية ومميزات الجمهور التنبؤية وتسمى بالتسويق السحابي التنبؤي. وتستخدم هذه الإضافة بيانات التيار الفورية لتحسين تفاعل العملاء عبر الإنترنت. أولًا، يحصل العملاء على درجة تنبؤية فريدة لهم، ويتم احتساب هذه النتيجة من عدة عوامل مختلفة، وتشمل كلًا من طول مدة سِجلّ التصفح، وما إذا كانوا قد قاموا بالنقر على رابط البريد الإلكتروني، وما إذا كانوا قد قاموا بشراء مادي، وكم أنفقوا، ومنذ متى قاموا بالشراء، وما إذا كانوا قد استجابوا لأي بريد إلكتروني أو حملة إعلانية. وبمجرد حصول العملاء على درجة، يتم تقسيمهم إلى مجموعات مختلفة؛ حيث يتم إعطاء هذه المجموعات خططًا وأهدافًا تسويقية مختلفة اعتمادًا على السلوكيات التنبؤية الخاصة بكل منهم، ويتم تحديث الدرجات والشرائح وتغييرها يوميًا مع إعطاء الشركات خارطة طريق أفضل لاستهداف وتحقيق الاستجابة المطلوبة. وتعد هذه الحلول التسويقية أكثر دقة كما أنها تتيح طرقًا أكثر تخصيصًا للشركات والتي يمكنها استيعاب أساليب الاحتفاظ بالعملاء.

أسئلة للمناقشة:

١- هل هناك مجالات في أي صناعة يكون فيها تيار البيانات غير ذي صلة؟

٢- بخلاف الاحتفاظ بالعملاء، ما هي المزايا الأخرى لاستخدام التحليلات التنبؤية؟

ماذا نتعلم من هذه الحالة العملية؟

تستطيع الشركات من خلال تحليل البيانات التي تم الحصول عليها في المكان والزمان المناسبَين، أن تقوم بتنبؤات وتتخذ قرارات بشأن المستهلكين بسرعة أكبر، مما يضمن تحقيق أهداف الأعمال، وجذب العملاء المناسبين، والاحتفاظ بهم وتعظيم قيمتهم لأقصى حد. ومن الجدير بالذكر أن البيانات التي تم الحصول عليها في الأسبوع الماضي، ليست مفيدة مثل البيانات التي لدى الشركات اليوم. كما أن استخدام البيانات ذات الصلة يجعل تحليلنا التنبؤي أكثر دقة وكفاءة.

Sources: Amodio, M. (2015). Salesforce adds predictive analytics to Marketing Cloud. Cloud Contact Center. http://www.cloudcon-tactcenterzone.com/topics/cloud contact-center/articles/413611-salesforce-adds-predictive analytics-marketing-cloud.htm (accessed July 2016). Davis, J. (2015). Salesforce adds new predictive analytics to Marketing Cloud. Information Week. http://www.informationweek.com/Big-data/Big-data analytics/salesforce-adds-new-predictive-analytics-to marketing-cloud/d/d-id/1323201 (accessed July 2016). Henschen, D. (2016). Salesforce reboots Wave Analytics, preps IoT cloud. ZD Net. http://www.zdnet.com/article/salesforce-reboots-wave-analytics-preps-iot-cloud/(accessed July 2016).

تطبيق القانون والحماية الإلكترونية:

توفـر تيارات البيانات الضخمة فرصًا ممتازة لتحسـين سُـبل منع الجريمـة، وتطبيق القانون، وتعزيز الأمان. فهي توفر إمكانات مذهلة في مجال تطبيقات الأمن التي يمكن بناؤها في الفضاء، مثل الوعي الظرفي الفوري، والمراقبة متعددة الوسـائط، وكشف الأمان الإلكتروني، والتصنت على المكالمات الهاتفية، ومراقبة الفيديو، والتعرف على الوجه (Zikopoulos وآخرون، ٢٠١٣).

مجال إنتاج الطاقة:

تزداد كمية البيانات التي تم جمعها بشـكل فوري عن طريق مرافق الطاقة، أضعافًا مضاعفة، بسـبب زيادة اسـتخدام العدادات الذكية. فقـد انتقلت من مرة كل شـهر لتصبح مرة كل ١٥ دقيقة (أو بشكل أكثر تكرارًا)، حيث تجمع قراءة العداد كميات كبيرة من البيانات القيمة لمرافق الطاقة. حيث ترسـل هـذه العدادات الذكية وغيرها من أجهزة الاستشـعار الموضوعة في جميع أنحاء الشـبكة الكهربائية، المعلومات إلى مراكز التحكم ليتم تحليلها فور وصولها، حيث تسـاعد مثل هذه التحليلات شركات المرافق لتحسـين قرارات سلاسـل التوريد الخاصة بهم (مثل: تعديل السـعة، وخيارات شـبكة التوزيع، والشراء أو البيع الفوري) وذلك من خلال الاعتماد على أنماط الاسـتخدام والطلب للمسـتهلكين، بالإضافـة إلى أن شركات المرافق يمكنها دمـج بيانات الطقس والظروف الطبيعية الأخرى في تحليلاتها لتحسـين توليـد الطاقة من مصادر بديلة (مثل: الرياح، والطاقـة الشمسـية) وتوقع أدق لمقـدار الطلب على الطاقـة في عمليات الحبيبـات الجغرافية المختلفة. كما أن هناك فوائد أخرى مشابهة تنطبق على المرافق الأخرى كالمياه والغاز الطبيعي.

الخدمات المالية:

تُعـد شركات الخدمـات المالية من بـين أبرز الأمثلة التي يمكـن أن توفر تحليلات تيـار البيانات الضخمـة لها قـرارات أسرع وأفضل، بالإضافة إلى الميزة التنافسـية التي تقدمها لهـا، وكذلك الرقابة

التنظيميـة. فهي توفر القدرة عـلى تحليل البيانات السريعة المتدفقة، والأحجام الضخمة من بيانات التداول عندما يكون وقت الاسـتجابة منخفض للغاية عبر الأسـواق والـدول، فضلًا عن تقديمها لميزة هائلة بشأن قرارات الشراء / البيع من النوع جزء من الثانية والتي يمكن أن تُترجم إلى مكاسب مالية كبيرة. وبالإضافة إلى قرارات الشراء / البيع المُثلى، فإن تحليل البيانات يمكنه أيضًا أن يسـاعد شركات الخدمات المالية في مراقبة التجارة بشكل فوري للكشف عن التزوير وغيره من الأنشطة غير القانونية.

علوم الصحة:

تسـتطيع الأجهزة الطبيـة حديثة العهد (مثل: أجهزة تخطيط القلـب والمعدات التي تقيس ضغط الدم، ومستوى الأكسجين في الدم، ومستوى السكر في الدم، ودرجة حرارة الجسم) أن تنتج بيانات تشخيصية / حساسة ذات قيمة عالية بمعدل سريع للغاية؛ حيث إن تسخير هذه البيانات وتحليلهـا فور إنتاجها يقدم فوائد - من النوع الذي نطلق عليه غالبًا «مسـألة حياة أو موت» - وذلك على عكس أي مجال آخر، فضلًا عن مساعدة شركات الرعاية الصحية لكي تصبح أكثر فعالية وكفاءة (مما يترتب عليه أن تصبح هذه الشركات أكثر تنافسـية وقادرة على تحقيق أرباح عالية)، وتعمل تحليلات التيار أيضًا على تحسين ظروف المرضى وإنقاذ الأرواح.

تقـوم العديد من أنظمة المستشـفيات في جميع أنحاء العالم بتطوير البنيـة التحتية للرعاية وأنظمة صحية مسـتقبلية. وتهدف هذه الأنظمة إلى الاستفادة بشكل كامل مما تقدمه التقنية، بل وأكثر من ذلك. إن اسـتخدام الأجهزة التي تولد بيانات عالية الدقة بمعدل سريع للغاية، إلى جانب أجهزة الحاسب فائقة السرعة التي يمكنها تحليل تيارات البيانات المتعددة بشكل متوازن، يـؤدي إلى زيادة فرص الحفاظ على سـلامة المرضى من خلال سرعة اكتشـاف الحالات الشـاذة. وتهـدف هذه الأنظمة إلى مسـاعدة صانعي القرارات الخاصة بالبشر عـلى اتخاذ قرارات أسرع وأفضل من خلال الاطلاع على عدد كبير من المعلومات بمجرد توفرها.

الحكومة:

تسعى الحكومات في جميع أنحاء العالم إلى إيجاد طرق ترفع من فعاليتها (من خلال الاستخدام الأمثـل للموارد المحدودة) وأكثر فاعلية (بتوفير الخدمـات التي يحتاجها الناس ويريدونها). ومع تحول ممارسات الحكومة الإلكترونية إلى الاتجاه السائد، إلى جانب الاستخدام واسع النطاق لوسائل التواصـل الاجتماعي والوصول إليهـا، فإن ذلك يوفر كميات ضخمة مـن البيانات (المهيكلة وغير المهيكلة) لتصبح تحت تصرف الجهات الحكومية. فالاسـتخدام السليم لتيارات البيانات الضخمة في الوقت المناسـب هو الذي يميز الجهات الاسـتباقية ذات الكفاءة العالية عن غيرها من الجهات

التي مازالت تســتخدم الطرق التقليدية للتفاعــل مع المواقف عند حدوثها. كما أن هناك طريقة أخرى تســتطيع الجهات الحكومية من خلالها الاستفادة من قدرات التحليلات الفورية وهي إدارة الكوارث الطبيعية مثل العواصف الثلجية، والأعاصير، والرياح الشديدة، وحرائق الغابات من خلال مراقبة تيار البيانات الصادر من أجهزة الرادار، وأجهزة الاستشعار، وأجهزة الكشف الذكية الأخرى. كما يمكنهم أيضًا اســتخدام أساليب مشــابهة لرصد جودة المياه، وجودة الهواء، وأنماط الاستهلاك واكتشــاف أي حالات مختلفة عن المعتاد قبل أن تصبح مشاكل كبيرة. وبالإضافة إلى ما سبق، فإن هناك مجالًا آخر تستخدم فيه الجهات الحكومية تحليلات التيار وهو مجال المرور، حيث تستخدم هــذه التحليلات في إدارة حركة المرور في المــدن المزدحمة؛ حيث يمكن للجهات الحكومية أن تغير إشــارات المرور الضوئية وحارات المرور لتخفيف الإزعاج الذي تســببه مشاكل الازدحام المروري، وذلك باســتخدام البيانات الصادرة عن كاميرات تدفق حركة المرور، وبيانات الـ GPS التي تصدر عن المركبات التجارية، وأجهزة استشعار حركة المرور الموجودة في الطرق.

أسئلة للمراجعة على القسم ٧-٨:

١- ما هي أكثر الصناعات فائدة لتحليلات التيار؟

٢- كيف يمكن استخدام تحليلات التيار في التجارة الإلكترونية؟

٣- بالإضافة إلى ما ذكر في هذا القسم، هل يمكنك التفكير في صناعات و / أو مجالات تطبيق أخرى يمكن استخدام تحليلات التيار فيها؟

٤- بالمقارنة مع التحليلات العادية، هل تعتقد أن تحليلات التيار ســتحتوي على حالات استخدام أكثر (أو أقل) في عصر تحليلات البيانات الضخمة؟ ولماذا؟

تلخيص لأهم نقاط الفصل:

- للبيانات الضخمة معانٍ مختلفة بالنسبة للأشخاص ذوي الخلفيات والاهتمامات المختلفة.
- تتخطــى البيانــات الضخمــة نطاق بيئة الأجهزة المســتخدمة و / أو إمكانيــات أدوات البرامج المستخدمة في جمعها وإدارتها ومعالجتها خلال فترة زمنية مقبولة.
- عــادةً ما يتم تعريف البيانات الضخمة من خلال ثلاث V: الحجم volume، والتنوع variety، والسرعة velocity.
- يعتبر MapReduce أســلوبًا لتقســيم المعالجة لملفات البيانــات الضخمة متعددة الهيكلة عبر تجمع كبير من الآلات.

- يعتبر هادوب إطار عمل مجاني لمعالجة وتخزين وتحليل كميات كبيرة من البيانات المقسمة غير المهيكلة.
- يعتبر Hive تخزين بيانات مبني على هادوب - يشبه إطار عمل تم تطويره أصلًا بواسطة Facebook.
- يعتبر Pig لغة استعلام مبنية على هادوب تم تطويرها بواسطة Yahoo.
- تعتبر NoSQL، والتي ترمز لـ Not Only SQL، نموذجًا جديدًا لتخزين ومعالجة كميات ضخمة من البيانات غير المهيكلة وشبه المهيكلة ومتعددة الهيكلة.
- يعتبر عالم البيانات Data scientist دورًا جديدًا أو وظيفةً جديدة ترتبط عادةً بالبيانات الضخمة أو علم البيانات.
- تعتبر كلٌ من البيانات الضخمة ومستودعات البيانات تقنيات متكاملة (وليست متنافسة) للتحليلات.
- كمجال جديد نسبيًا، يتطور مورد البيانات الضخمة بشكل سريع جدًا.
- تعتبر تحليلات التيار Stream analytics مصطلحًا يُستخدم عادةً لاستخراج المعلومات القابلة للتطبيق من مصادر البيانات المتدفقة باستمرار.
- تعمل التحليلات الدائمة على تقييم كل مشاهدة داخلة في مقابل جميع المشاهدات السابقة.
- تعتبر معالجة الأحداث الحرجة طريقة لالتقاط وتتبع وتحليل تيارات البيانات لاستكشاف أحداث معينة (بخلاف الأحداث العادية) والتي تستحق هذا المجهود.
- يعرف التنقيب في تيار البيانات، كتقنية تمكينية لتحليلات التيار، على أنه عملية لاستخراج الأنماط الجديدة وهياكل المعرفة من سجلات البيانات المستمرة والسريعة.

مصطلحات أساسية:

البيانات الضخمة.	التنقيب في تيار البيانات.	Hive.
تحليلات البيانات الضخمة.	Hadoop.	MapReduce.
معالجة الأحداث الحرجة.	ملف هادوب المقسم.	NoSQL.
نظام (HDFS).	التحليلات الدائمة.	Pig.
تحليلات التيار.	علماء البيانات.	

أسئلة للمناقشة:

١- ما هي البيانات الضخمة؟ ولماذا تعتبر مهمّة؟ ومن أين تأتي؟

٢- في رأيك، ما هو مستقبل البيانات الضخمة؟ وهل ستفقد شعبيتها إلى شيءٍ آخر؟ إذا كانت الإجابة بنعم، فماذا سيكون ذلك الشيء؟

٣- ما هي تحليلات البيانات الضخمة؟ وكيف تختلف عن التحليلات العادية؟

٤- ما هي العوامل المهمة لنجاح تحليلات البيانات الضخمة؟

٥- ما هي التحديات الكبيرة التي يجب مراعاتها عند اعتماد تنفيذ تحليلات البيانات الضخمة؟

٦- ما هي مشاكل الأعمال الشائعة التي تعالجها تحليلات البيانات الضخمة؟

٧- في عصر البيانات الضخمة، هل نعتبر بصدد نهاية عصر مستودعات البيانات؟ ولماذا؟

٨- ما هي حالات استخدام البيانات الضخمة / هادوب ومستودعات البيانات / RDBMS؟

٩- ما هي تحليلات التيار؟ وكيف تختلف عن التحليلات العادية؟

١٠- ما هي أكثر الصناعات المجدية لتحليلات التيار؟ وما هو القاسم المشترك بين تلك الصناعات؟

١١- بالمقارنة مع التحليلات العادية، هل تعتقد أن تحليلات التيار ستحتوي على حالات استخدام أكثر (أو أقل) في عصر تحليلات البيانات الضخمة؟ ولماذا؟

تمارين - شبكة جامعة تيراداتا (TUN) وغيرها من التمارين اليدوية:

١- اذهب إلى teradatauniversitynetwork.com، وابحث عن دراسات الحالة. قم بقراءة الحالات والأوراق البيضاء التي تتحدث عن تحليلات البيانات الضخمة. ما هي الفكرة الرئيسة المشتركة بين دراسات الحالة تلك؟

٢- في teradatauniversitynetwork.com، قم بإيجاد الأوراق البيضاء لـ تحليلات SAS المرئية، ودراسات الحالة، والتدريبات اليدوية. قم بإجراء تمارين التحليلات المرئية على مجموعات البيانات الضخمة ثم قم بإعداد تقرير لمناقشة النتائج التي توصلت إليها.

٣- في teradatauniversitynetwork.com، اذهب إلى صفحة Sports Analytics. قم بإيجاد تطبيقات البيانات الضخمة في الرياضة. لخّص النتائج التي حصلت عليها.

٤- اذهب إلى teradatauniversitynetwork.com، وابحث عن مقاطع فيديو BSI التي تتكلم عن البيانات الضخمة. راجع هذه المقاطع، ثم أجب عن أسئلة الحالة المتعلقة بها.

٥- اذهب إلى موقعَي teradata.com وasterdata.com أو أحدهما. قم بإيجاد على الأقل ثلاث دراسات حالة عملاء على البيانات الضخمة، واكتب تقريرًا تناقش فيه أوجه الشبه والاختلاف بين هذه الحالات.

٦- اذهب إلى IBM.com. قم بإيجاد على الأقل ثلاث دراسات حالة عملاء على البيانات الضخمة، واكتب تقريرًا تناقش فيه أوجه الشبه والاختلاف بين هذه الحالات.

٧- اذهب إلى claudera.com. قم بإيجاد على الأقل ثلاث دراسات حالة عملاء على تنفيذ هادوب، واكتب تقريرًا تناقش فيه أوجه الشبه والاختلاف بين هذه الحالات.

٨- اذهب إلى mapr.com. قم بإيجاد على الأقل ثلاث دراسات حالة عملاء على تنفيذ هادوب، واكتب تقريرًا تناقش فيه أوجه الشبه والاختلاف بين هذه الحالات.

٩- اذهب إلى hortonworks.com. قم بإيجاد على الأقل ثلاث دراسات حالة عملاء على تنفيذ هادوب، واكتب تقريرًا تناقش فيه أوجه الشبه والاختلاف بين هذه الحالات.

١٠- اذهب إلى marklogic.com. قم بإيجاد على الأقل ثلاث دراسات حالة عملاء على تنفيذ هادوب، واكتب تقريرًا تناقش فيه أوجه الشبه والاختلاف بين هذه الحالات.

١١- اذهب إلى youtube.com. ابحث عن مقاطع فيديو عن حوسبة البيانات الضخمة. شاهد اثنين منهم على الأقل. لخِّص النتائج التي حصلت عليها.

١٢- اذهب إلى google.com/scholar وابحث عن مقالات عن تحليلات التيار. قم بقراءة ثلاث مقالات منها على الأقل. لخِّص النتائج الخاصة بك.

١٣- ادخل على google.com/scholar، وابحث عن مقالات عن التنقيب في تيار البيانات. قم بقراءة ثلاث مقالات منها على الأقل. لخِّص النتائج الخاصة بك.

١٤- ادخل على google.com/scholar، وابحث عن مقالات تتكلم عن البيانات الضخمة في مقابل مستودعات البيانات. قم بقراءة خمس مقالات منها على الأقل. لخِّص النتائج الخاصة بك.

المراجع:

- Adshead, A. (2014). Data set to grow 10-fold by 2020 as Internet of Things takes off. http://www.computerweekly.com/news/2240217788/Data-set-to-grow-10-fold-by-2020-as-internet-of-things-takes-off (accessed September 2016).
- Amodio, M. (2015). Salesforce adds predictive analytics to Marketing Cloud. Cloud Contact Center. cloudcontactcenter-zone.com/topics/cloud-contact-center/articles/413611-salesforce-adds-predictive-analytics-marketing-cloud.htm (accessed August 2016).
- Asamoah, D., Sharda, R., Zadeh, A., & Kalgotra, P. (2016). Preparing a Big Data analytics professional: A pedagogic experience. In DSI 2016 Conference, Austin, TX.
- Asamoah, D., & Sharda, R. (2015). Adapting CRISP-DM process for social network analytics: Application to healthcare. In AMCIS 2015 Proceedings. aisel.aisnet.org/amcis2015/bizAnalytics/GeneralPresentations/33/(accessed July 2016).
- Awadallah, A., & Graham, D. (2012). Hadoop and the data warehouse: When to use which. teradata.com/white-papers/Hadoop-and-the-Data-Warehouse-When-to-Use-Which (accessed August 2016).
- Broniatowski, D. A., Paul, M. J., & Dredze, M. (2013). National and local influenza surveillance through Twitter: An analysis of the 2012-2013 influenza epidemic. PloS One, 8(12), e83672.
- Cisco. (2016). The zettabyte era: Trends and analysis. cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.pdf (accessed August 2016).
- DataStax. Customer case studies. datastax.com/resources/case studies/eBay (accessed July 2016).
- Davis, J. (2015). Salesforce adds new predictive analytics to Marketing Cloud. Information Week. informationweek.com/Big-data/Big-data-analytics/salesforce-adds-new-predictive-analytics-to-marketing-cloud/d/d-id/1323201 (accessed August 2016).
- Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified data processing on large clusters. research.google.com/archive/mapreduce.html (accessed August 2016).
- Delen, D., Kletke, M., & Kim, J. (2005). A scalable classification algorithm for very large datasets. Journal of Information and Knowledge Management, 4(2), 83-94.
- Dillow, C. (2016). What happens when you combine artificial intelligence and satellite imagery. fortune.com/2016/03/30/facebook-ai-satellite-imagery/(accessed July 2016).
- Ekster, G. (2015). Driving investment performance with alternative data. integrity-research.com/wp-content/uploads/2015/11/Driving-Investment-Performance-With-Alternative-Data.pdf (accessed July 2016).
- Henschen, D. (2016). Salesforce reboots Wave Analytics, preps IoT cloud. ZD Net. zdnet.com/article/salesforce-reboots- wave-analytics-preps-iot-cloud/(accessed August 2016).
- Higginbotham, S. (2012). As data gets Bigger, what comes after a yottabyte? gigaom.com/2012/10/30/as-data-gets- Bigger-what-comes-after-a-yottabyte (accessed August 2016).

- Hope, B. (2015). Provider of personal finance tools tracks bank cards, sells data to investors. Wall Street Journal.wsj.com/articles/provider-of-personal-finance-tools-tracks-bank-cards-sells-data-to-investors-1438914620 (accessed July 2016).
- Jonas, J. (2007). Streaming analytics vs. perpetual analytics (Advantages of Windowless Thinking). jeffjonas.typepad.com/jeff_jonas/2007/04/streaming_analy.html (accessed August 2016).
- Kalgotra, P., & Sharda, R. (2016). Rural versus urban comorbidity networks. Working Paper, Center for Health Systems and Innovation, Oklahoma State University.
- Kelly, L. (2012). Big data: Hadoop, business analytics, and beyond. wikibon.org/wiki/v/Big_Data:_Hadoop,_Business_Analytics_and_Beyond (accessed August 2016).
- Lampitt, A. (2012). Big data visualization: A Big deal for eBay. infoworld.com/d/Big-data/Big-data-visualization-Big-dealebay-208589 (accessed August 2016).
- MarkLogic. (2012). Top 5 investment bank achieves single source of truth. marklogic.com/resources/top-5-derivatives-trading-bank-achieves-single-source-of-truth (accessed July 2016).
- Moran, P. A. (1950). Notes on continuous stochastic phenomena. biometrika, 17-23.
- Orbital Insight. World Oil Storage Index. orbitalinsight.com/solutions/world-oil-storage-index/ (accessed July 2016).
- Russom, P. (2013). Busting 10 myths about Hadoop: The Big Data explosion. TDWI's Best of Business Intelligence, 10, 45-46.
- Sarasohn-Kahn, J. (2008). The wisdom of patients: Health care meets online social media. Oakland, CA: California HealthCare Foundation.
- Shaw, C. (2016). Satellite companies moving markets. quandl.com/blog/alternative-data-satellite-companies (accessed July 2016).
- Steiner, C. (2009). Sky high tips for crop traders (accessed July 2016).
- St Louis, C., & Zorlu, G. (2012). Can Twitter predict disease outbreaks? BMJ, 344.
- Tableau white paper. (2012). 7 Tips to succeed with Big Data in 2013. cdnlarge.tableausoftware.com/sites/default/files/whitepapers/7-tips-to-succeed-with-Big-data-in-2013.pdf (accessed August 2016).
- Thusoo, A., Shao, Z., & Anthony, S. (2010). Data warehousing and analytics infrastructure at Facebook. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data (p. 1013).
- Turner, M. (2015). This is the future of investing, and you probably can't afford it. businessinsider.com/hedge-funds-are-analysing-data-to-get-an-edge-2015-8 (accessed July 2016).
- Watson, H. (2012). The requirements for being an analytics-based organization. Business Intelligence Journal, 17(2), 42-44.
- Watson, H., Sharda, R., & Schrader, D. (2012). Big Data and how to teach it. Workshop at AMCIS, Seattle, WA.
- White, C. (2012). MapReduce and the data scientist. Teradata Aster White Paper. teradata.com/white-paper/MapReduce-and-the-Data-Scientist (accessed August 2016).

- Wikipedia.com. "Petabyte."en.wikipedia.org/wiki/Petabyte (accessed August 2016).
- Zadeh, A. H., Zolbanin, H. M., Sharda, R., & Delen, D. (2015). Social media for nowcasting the flu activity: Spatial-temporal and text analysis. Business Analytics Congress, Pre-ICIS Conference, Fort Worth, TX.
- Zikopoulos, P., DeRoos, D., Parasuraman, K., Deutsch, T., Corrigan, D., & Giles, J. (2013). Harness the power of Big Data. New York: McGraw-Hill.

الفصل الثامن
الاتجاهات المستقبلية والخصوصية
والاعتبارات الإدارية في التحليلات

أهداف التعلم:

- استكشاف بعض التقنيات حديثة الظهور التي قد توثر على التحليلات، وذكاء الأعمال، ودعم القرار.
- وصــف الظاهرة حديثة الظهور المعروفة باســم إنترنت الأشــياء (IoT)، والتطبيقات المحتملة، والنظام الاقتصادي لظاهرة (إنترنت الأشياء).
- وصف الاستخدام الجاري والمستقبلي للحوسبة السحابية في تحليلات الأعمال.
- وصف كيفية مساعدة التحليلات المبنية على الجغرافية المكانية والموقع للمنظمات.
- وصف التأثيرات التنظيمية لتطبيقات التحليلات.
- سرد ووصف القضايا الأخلاقية والقانونية الرئيسة لتنفيذ التحليلات.
- تحديد الخصائص الرئيسة لمحترفي علم البيانات الناجح.

يقدم هذا الفصل عدة تقنيات حديثة الظهور والتي من المحتمل أن تكون ذات تأثيرات جوهرية على تطوير واستخدام تطبيقات ذكاء الأعمال. وفي مجال حركي كالتحليلات، فإن المصطلحات أيضًا تتطــور وتتداخــل. وكما ســبق ذكره، فيمكننا الرجــوع إلى هذه التقنيات مثــل ذكاء الأعمال، أو التحليــلات، أو علم البيانات، أو تعلم الآلة، أو الذكاء الاصطناعي (AI)، أو الحوســبة المعرفية، أو البيانات الضخمة، أو العديد من العناوين الأخرى. وليس هدفنا هو التركيز على الفروق الطفيفة بــين كل منهــا، ولكن الهدف هوالتعامل معهم جميعًا ككوكبة كبــيرة واحدة. إننا نركز على بعض الاتجاهــات التي تحققت بالفعل وعــلى البعض الآخر المتوقع أن يؤثر بصورة أكبر على التحليلات. ودائمًــا ما يعد اســتخدام كرة كريســتال اقتراحًا منذرًا بالخطر، ولكن في هــذا الفصل نقدم تحليلًا لبعض المجالات النامية. كما نقدم ونشرح بعض التقنيات حديثة الظهور وكذلك نستكشف بعض تطبيقاتهــا المعاصرة، ثم نناقش بعد ذلــك التأثيرات التنظيمية والشــخصية والقانونية والأخلاقية والاجتماعية لأنظمة الدعم التحليلي والقضايا التي يجب أن يهتم بها مديرو ومحترفو التحليلات.

٨-١ مقال افتتاحي: تحليل بيانات الاستشــعار يســاعد ســيمنس (Siemens) على تجنب أعطال القطارات:

تعتــبر شركة Siemens أكــبر منتج في العالم لتقنيات كفاءة الطاقــة وتوفير الموارد. يقع المقر الرئيسي للشركة في برلين بألمانيا، بعائدات ٩٣ مليار دولار سنويًا. تنتج Siemens قطارات متنوعة ومكونات بنية تحتية تشبه أنظمة التحكم وأنظمة الطاقة.

يعتقــد فريق خدمات نقــل البيانات التابع لشركة Siemens أن تحليــلات البيانات الضخمة وإنترنت الأشــياء (IoT) يســتطيع أن يمكنهم مــن توقع أخطاء المكونات لأســابيع قادمة. إنهم يستكشــفون هذه الأســاليب للتأكد من عدم ترك أي قطار محجوز على المسارات بسبب أعطال فنيــة غير متوقعة. وتريــد Siemens الانتقال مــن الصيانة التفاعلية (بعد الحــادث) والصيانة الوقائية (بعمليات التفتيش المنتظمة) إلى الصيانة التنبؤية للقطارات.

تقيس أجهزة الاستشــعار المتصلة بمكونات قطاراتهــا الموقف الحالي للمكونات، حيث تقوم شركة Siemens بجمع وتحليل بيانات جهاز الاستشعار بشكل فوري تقريبًا. وإذا تم العثور على أي شذوذ في البيانات، فهذا يشير إلى احتمال عطل أحد المكونات. وبالتالي، يمكن اتخاذ تدابير وقائية وفقًا لذلك.

وفقًا لمدونة تيراداتا، يعتمد مهندسو Siemens على بيانات عشرات الآلاف من أجهزة الاستشعار. تذهب كل من بيانات القطارات والسكك الحديدية، وبيانات عمليات الإصلاح، وبيانات الطقس، وبيانات سلاسل التوريد، إلى بنية البيانات الموحدة تيراداتا التابعة لشركة Siemens للاستفادة من Hadoop، وTeradata Aster، ومســتودع بيانات تيراداتا. وصرح Gerhard Kress مدير فريق خدمات نقل البيانات بشركة Siemens قائلًا: « لا يمكننا أن نفعل ما نفعله بناءً على بنية مختلفة لأن أحجام البيانات لدينا كبيرة جدًا. ولذا؛ فمثلًا بالنسبة لأسطول سيارات من أوروبا، قمنا بجمع كل بيانات أجهزة الاستشــعار، وكانت حوالي ١٠٠ مليار سطرًا جدوليًا. وإذا كنت ترغب في تنفيذ خوارزمية تعلم الآلة على ذلك فهي لا تعمل على شيء ليس موازياً بشكل كبير».

يســمح تعلم الآلة باســتخدام بيانــات جهاز الاستشــعار لعلماء ومهنــدسي البيانات بشركة Siemens من الكشــف السريع عن الإيجابيات الخاطئة (التنبــؤ بعطل لا يحدث فعلًا) وإعطاء تنبــؤ واضح بأعطال الجــزء الفعلي. ولأن الإنذارات الكاذبة تكون أكــثر من الإنذارات الحقيقية، تقــوم المنظمة بالنظر في أوامر العمل، والأرقام التسلســلية، وتاريخ بيانات القطارات والخدمات، ومعلومات التشــخيص، وبيانات جهاز الاستشــعار، وعمليات الإصلاح، وبيانات سلاســل التوريد للمســاعدة في تحديد وحل أعطــال الجزء الحقيقي. وتســتطيع Siemens أن تحدد بدقة أيهما أكثر عرضة للعطل، هل هو القطار فائق السرعة بين موســكو وســانت بتسبرج في الشتاء القارس أم القطار فائق السرعة في إسبانيا الحارة في أيام الصيف، وذلك عن طريق دمج بيانات الطقس.

يســتخدم مشغل القطار الأســباني RENFE المكونات الأساســية لقطار Siemens عالي السرعة، Valero E، الخاضع للمراقبة المستمرة من جانب Siemens. فإذا كانت أنماط بيانات جهاز الاستشعار التي تم جمعها غير طبيعية، يتم إرسال فريق للتحقق من هذه المكونات، وبالتالي يمنع عطل القطار على المسارات. وكنتيجة لذلك، لم يتأخر بشكل ملحوظ سوى قطار واحد فقط من أصل ٢٣٠٠ قطار. وتزعم Siemens أيضًا أن المصداقية المرتفعة لقطاراتها قد ساعدت الكثير من مشغلي القطارات على تحسين أدائهم فيما يتعلق بالوقت المناسب. فعلى سبيل المثال، تعمل قطارات بانكوك بنسبة تأخير ١٪ فقط. وتقدِّر Siemens كذلك أن التشــغيلات التي يمكن الاعتماد عليها بشكل كبير بين برشلونة ومدريد قد أدت إلى زيادة معنوية في عدد ركاب القطارات وتقليص حركة الملاحة الجوية.

وهكذا، تكون النمذجة التنبؤية قد منحت Siemens فرصة خدمية جديدة. لقد تم تأسيسها الآن كمــورد رئيــس لخدمة تتبع أعطال القطارات. وســيصبح بيع هذه الخدمــات التنبؤية أكثر أهمية من مجرد المعدات الأصلية.

ما هي الدروس التي يمكننا تعلمها من هذه المقالة الافتتاحية؟

تقــود Siemens تطوير الخدمات الإضافية والقيمة للمنتجــات الصناعية الضخمة القادمة من الصناعات الكبيرة نسبيًّا؛ إذ يمكن للشركة إنشاء صورة أفضل لطريقة أداء منتجاتها على أرض الواقع، وذلك من خلال استعمال أجهزة الاستشعار التي تولد كميات كبيرة ومتنوعة من البيانات ودمج تلك البيانات مع مصادر البيانات الأخرى مثل الطقس. وعلاوةً على ذلك، فإن تحليل مثل هذه البيانات قد يساعد العميل على أداء الصيانة فقط عندما يلزم ذلك وليس وفقًا لجدول زمني. ولقد أصبح بيع مثل هذه الخدمات التحليلية محور اهتمام أســاسي جديد بالنسبة لجميع المنتجين الروّاد للمعدات الصناعية مثل Siemens وGeneral Electric. كما يعتبر بيع خدمات أداء التحليلات على المنتجات والصيانة التنبؤية أو الإصلاحات مثالًا ممتازًا لإنشاء فرص تسويقية جديدة للمنتجات القائمة.

أسئلة للمناقشة:

١- في مجال المعدات الصناعية مثل القطارات، ما هي المَعْلَمات التي يمكن قياسها بانتظام لتقدير الأداء الحالي للمعدات والإصلاحات اللازمة مستقبلًا؟

٢- كيف يمكن أن نستفيد من بيانات الطقس في تحليل حالة معدات القطارات؟

٣- قدِّر كمية البيانات التي تحتاج لجمعها في شهر واحد باستخدام مثلًا ١٠٠٠ جهاز استشعار على قطار واحد، علمًا بأن كل جهاز استشعار يعطي ١ كيلوبايت من البيانات في الثانية الواحدة.

٤- ما هو تصورك لتخزين مثل هذه المجموعات من البيانات؟

٨-٢ إنترنت الأشياء:

قدمت المقالة الافتتاحية مجالًا يواجه نموًا كبيرًا في الوقت الراهن. يعتبر إنترنت الأشياء (IoT) ظاهرة لربط العالم المادي بالإنترنت، بخلاف إنترنت الأشخاص الذي يربط البشر ببعضهم البعض من خلال التقنية. ففي إنترنت الأشياء، تتصل الأجهزة المادية بأجهزة الاستشعار التي تجمع بيانات عن الجهاز مثل التشغيل والموقع والحالة الراهنة. تتم معالجة هذه البيانات باستخدام أساليب تحليلية متنوعة لمراقبة الجهاز عن بُعد من مكتب مركزي أو للتنبؤ بأي أخطاء قد تحدث للجهاز. ولعل المثال الأشهر على إنترنت الأشياء هو السيارة ذاتية القيادة. ولكي تقود نفسها بنفسها، تحتاج السيارة وجود أجهزة استشعار كافية لمراقبة الموقف تلقائيًا من حولها وتتخذ الإجراءات الملائمة لضبط أي إعداد لازم، بما في ذلك سرعة السيارة واتجاهها وما إلى ذلك. ومن الأمثلة الشائعة الأخرى على إنترنت الأشياء هو جهاز متتبع اللياقة الذي يسمح للمستخدم بتتبع أنشطته الجسدية كالمشي والجري والنوم. وتعد شركة Smartbin مثالًا آخر لتوضيح ظاهرة إنترنت الأشياء. فقد قامت Smartbin بتطوير حاويات القمامة ذات أجهزة استشعار للكشف عن مستويات الامتلاء؛ حيث يقوم جهاز الاستشعار بإخطار شركة جمع القمامة تلقائيًا بأن الحاويات قد امتلأت تقريبًا وتحتاج للتفريغ. وبالطبع، يعد المثال الأكثر شيوعًا الذي يضربه الناس لتوضيح إنترنت الأشياء هو فكرة طلب الثلاجة للحليب تلقائيًا عندما تكتشف أن الحليب قد نفد. وقدمت شركة كلوروكس فلتر Brita الجديد المزود بوعاء به Wi-Fi يستطيع طلب فلاتر المياه بنفسه عند اكتشاف وقت التغيير. في كل هذه الأمثلة، لا يضطر البشر إلى التواصل مع بعضهم البعض، ولا حتى مع أي آلة في كثير من الحالات؛ حيث تستطيع الآلات أن تتكلم. وهذا هو سبب استخدام مصطلح إنترنت الأشياء.

وفقًا لبحث Juniper (٢٠١٦)، فبالإضافة إلى أجهزة التابلت والهواتف الذكية والحواسيب الشخصية، سيتم ربط أكثر من ٣٨ مليار شيء بالإنترنت وذلك بحلول عام ٢٠٢٠. وفيما يلي بعض الأسباب للنمو الكبير في إنترنت الأشياء:

١- كون الأجهزة أصغر حجمًا وأرخص سعرًا وأكثر قوةً: لقد انخفضت تكاليف المشغلات وأجهزة الاستشعار بشكل كبير في العشر سنوات المنصرمة، مما أدى بشكل عام لإنتاج جهاز استشعار أرخص كثيرًا.

٢- إمكانية التنقل الرخيصة: وقد انخفضت كذلك تكاليف معالجة البيانات والسعة التنزيلية وأجهزة الجوال بنسبة ٩٧٪ منذ العقد الماضي.

Source: Adapted from teradata.com (2016). The Internet of trains. http:// www.teradata.com/resources/case-studies (accessed August 2016); theit.org. (2016). Siemens tracks Big Data for trains that keep on running. http:// eandt.theiet.org/magazine/2016/07/siemens-Big-data-trains.cfm (accessed August 2016); Siemens.com (2016). About Siemens. http:// www.siemens.com/about/en/(accessed August 2016).

٣- توافر أدوات ذكاء الأعمال: تعرض الآن الكثير والكثير من الشركات أدوات ذكاء الأعمال الخاصة بها سواء على أجهزتها أو على السحابة بمعدلات أقل. وتكون أدوات البيانات الضخمة وذكاء الأعمال متاحة على نطاق واسع وبصورة متطورة للغاية.

٤- تظهر يوميًّا بشكل افتراضي حالات استخدام جديدة ومثيرة.

كما ينبغي علينا ملاحظة وجود بعض التعارض حول استخدام مصطلح إنترنت الأشياء. وأيضًا يطلق عليه البعض مصطلح ويب الأشياء Web of Things، ويسميه البعض الآخر إنترنت الأنظمة لأنه في العديد من الطرق سيكون هناك توليفة من الأنظمة التي تتواصل فيما بينها. عمومًا، سنشير في هذا الفصل لهذه الظاهرة باسم إنترنت الأشياء (IoT) لتحقيق الاتساق.

وتختلف التقديرات على نطاق واسع فيما يخص نمو صناعة إنترنت الأشياء، في حين سيتم إنفاق مبلغ يتراوح ما بين ٦ تريليون و١١ تريليون دولار تقريبًا على حلول إنترنت الأشياء بقدوم عام ٢٠٢٠، والتي تعد واحدة من أسرع قطاعات تقنية المعلومات نموًا بشكل عام وعنصرًا أساسيًّا في صناعة التحليلات. وتعطي الحالتان العمليتان (٨-١) و(٨-٢) مثالين لتطبيقات إنترنت الأشياء في صناعات مختلفة باستخدام منصتين تقنيتين مختلفتين.

حالة عملية ٨-١

شركة زوارق SilverHook تستخدم تحليل البيانات الفوري لإخطار المتسابقين والمشجعين

تعتبر زوارق SilverHook إحدى أسرع القوارب في العالم والأكفأ في استهلاك الطاقة والتي تصنعها شركة تحمل نفس الاسم. وتستخدم زوارق SilverHook في جميع سباقات الزوارق حول العالم. تواجه القوارب المنطلقة بسرعات عالية ضد المياه المالحة خطر عطل المعدات. ويعتمد المتسابقون على بيانات القياس عن بُعد من قواربهم لاتخاذ قرارات تتعلق بالسلامة والتخطيط. وفي بعض المواقف، يتم أيضًا جمع بيانات القياس الحيوية للمتسابقين. وبالتالي، يستقبل المتسابقون الكثير من المعلومات، مما يُصعِّب تتبع كل البيانات. كما لاحظت SilverHook أن جمهور سباقات الزوارق كانوا يواجهون صعوبة لمتابعة المتسابقين في المحيط، كلما انطلقت القوارب بسرعة، مما قلل من تعلق الجمهور باللعبة. وقد جمعت SilverHook البيانات من ٨٠ جهاز استشعار ولكن لم تكن هناك طريقة سهلة للحصول على أفكار لتحسين قدرات اتخاذ القرار واتخاذ قرارات السلامة وتحسين تجربة الجمهور.

استخدم فريق SilverHook منصة IBM Bluemix كخدمة PaaS (انظر القسم التالي) لاستخدام حلول تحليلات IBM SPSS وتوصيل أفكار مفهومة للمستخدمين والجماهير. وقد تمكنت SilverHook عن طريق PaaS من إنشاء تطبيقات، وسحب بيانات في السحابة، وإجراء التحليلات المطلوبة على تلك البيانات. واستخدم IBM Bluemix، بالاشتراك مع شركة Virtual Eye، قاعدة إنترنت الأشياء لإرسال بيانات الاستشعار المتولدة من أجهزة الاستشعار ونظام تحديد المواقع العالمي (GPS) إلى السحابة. وقد تم إجراء تحليل على بيانات جهاز الاستشعار باستخدام أدوات التحليل الموجودة في IBM Bluemix. كما قامت Virtual Eye بتوصيل تنبيهات المعلومات الفنية للفريق المتسابق والتمثيل الفوري للسباق إلى الجماهير. والآن يستطيع المتسابقون الحصول على رؤية فورية عن السباق، لتحسين قراراتهم وقدرتهم التنافسية.

أسئلة للمناقشة:

١- ما هو نوع المعلومات المهمة للمتسابقين والتي يمكن أن تنشئها أجهزة الاستشعار على قارب السباق؟ وماذا عن الجماهير؟

٢- ما هي الرياضات الأخرى التي يمكن أن تستفيد من تقنيات مشابهة؟

٣- ما هي التحديات التقنية التي يمكن أن تواجهها في إنشاء مثل هذه الأنظمة؟

Sources: IBMBigdatahub.com. (2015). SilverHook Powerboats: Tracking fast-moving boats in real time. http://www.ibmBigda-tahub.com/blog/silverhook-powerboats-tracking-fast-moving-powerboats-real-time (accessed August 2016); IBM.com (2015). Case study: SilverHook Powerboats. http://www.ibm.com/cloud-computing/case studies/silverhook-powerboats.html (accessed August 2016).

حالة عملية ٨-٢

شركة Rockwell Automation تراقب أصول استكشاف النفط والغاز غالية الثمن

تعتبر شركة Rockwell Automation إحدى أكبر شركات توريد حلول التشغيل الآلي الصناعي والمعلومات الصناعية في العالم؛ إذ ينتشر عملاؤها في أكثر من ٨٠ دولة حول العالم، ويعمل بها حوالي ٢٢٥٠٠ موظف. وترتكز الشركة في عملها على عدة مجالات، يعد أحدها مساعدة شركات النفط والغاز في الاستكشاف. ومن أمثلة ذلك شركة Hilcorp Energy، وهي شركة تقوم بالتنقيب عن النفط في ألاسكا، وهي تستخدم معدات غالية

الثمن جدًا في الحفر والاستخراج وتكرير النفط، ويمكن لعطل واحد في المعدات أن يكلف شركة النفط والغاز فاقد إنتاج بما يتراوح من ١٠٠ ألف إلى ٣٠٠ ألف دولار في اليوم. ويتطلب التعامل مع هذه المشكلة وجود تقنية يمكنها مراقبة حالة مثل هذه المعدات عن بُعد، كما يمكنها التنبؤ بالمشكلات التي من المحتمل أن تحدث بشكل جيد في وقت مبكر.

اقتنصت Rockwell Automation هذه الفرصة لتوسيع أعمالها في صناعات النفط والغاز من خلال جمع البيانات من مواقع الاستكشاف وتحليلها لتحسين الكفاءة وقيادة الأداء بشكل أفضل. وتحصل الشركة على رؤيتها من Connected Enterprise باستخدام إنترنت الأشياء التابعة لمايكروسوفت (Microsoft) لتوفير مراقبة ودعم معدات النفط والغاز التي تحتل مكانًا في مناطق بعيدة. وتوفر Rockwell Automation الآن حلولًا للتنبؤ الفوري بأعطال المعدات على طول سلاسل توريد البترول، ومراقبة صحتها وأدائها، والمساعدة في منع عطلها في المستقبل. وتتركز هذه الحلول في المجالات التالية:

١- الحفر: تمتلك شركة Hilcorp Energy معدات الضخ الخاصة بها المتمركزة في ألاسكا، حيث تقوم بالتنقيب عن النفط ٢٤ ساعة في اليوم، ويمكن لعطل واحد في المعدات أن يكلف الشركة الكثير من المال. وقامت Rockwell Automation بربط المشغلات الكهربائية المتغيرة لمعدات الضخ بالسحابة (انظر القسم التالي)، بحيث يمكن التحكم في هذه الآلات على بعد آلاف الأميال من غرفة التحكم في أوهايو؛ إذ تقوم أجهزة الاستشعار بالتقاط البيانات، وتمريرها إلى Microsoft Azure Cloud عن طريق بوابة التحكم الخاصة بشركة Rockwell، ثم تصل هذه البيانات إلى المهندسين التابعين لـ Hilcorp Energy من خلال لوحات معلومات رقمية، فتقوم لوحات المعلومات هذه بعرض معلومات عن الضغط ودرجة الحرارة ومعدل التدفق وعشرات المَعْلَمات الأخرى بشكل آني مما يساعد المهندسين على مراقبة صحة وأداء المعدات. كما تعرض لوحات المعلومات هذه تنبيهات عن أي مشكلات محتملة. فعندما تعطلت إحدى قطع معدات الضخ الخاصة بشركة Hilcorp، تم تحديد العطل وتعقبه وإصلاحه في أقل من ساعة، مما أدى لتوفير ٦ ساعات من تعقب العطل والتكلفة الناجمة عن فاقد الإنتاج.

٢- بناء مضخات غاز أذكى: في هذه الأيام، تقوم بعض شاحنات النقل باستخدام الغاز الطبيعي السائل كوقود. وتقوم شركات النفط بتحديث محطات تعبئة الوقود الخاصة بها لدمج مضخات الغاز الطبيعي السائل. كما قامت شركة Rockwell Automation بتثبيت أجهزة استشعار ومشغلات تردد متغير على هذه الشاحنات لجمع بيانات فورية عن تشغيل المعدات ومخزون الوقود ومعدل الاستهلاك. وتتم إحالة هذه البيانات إلى منصة

السحابة الخاصة بشركة Rockwell، والتي تقوم بدورها بإنشاء لوحات معلومات تفاعلية وإعداد تقارير لوصف هذه التفاصيل باســتخدام Microsoft Azure ومن ثم إرسالها إلى المستفيدين الملائمين، مما يعطي انطباعًا مريحًا للمستفيدين عن سلامة أصول رأسمالهم.

لقد عملت حلــول Connected Enterprise المقدمة مــن شركة Rockwell على زيــادة سرعة نمو العديد مــن شركات النفط والغاز مثل شركــة Hilcorp Energy من خلال وضع بيانات العمليات الخاصة بها على منصة الســحابة ومســاعدتها على تقليل التكلفة الناجمة عن أوقات الأعطال والصيانة. ونتج عن هذا ظهور فرص عمل جديدة للمصنعين الشجعان مثل Rockwell Automation.

أسئلة للمناقشة:

١- مــا هو نوع المعلومات التي مــن المرجح أن يتم جمعها من قِبَل منصة التنقيب عن النفط والغاز؟

٢- هل يتناســب هذا التطبيق مع الثلاث V's (الحجــم والتنوع والسرعة) الخاصة بالبيانات الضخمة؟ لماذا نعم أو لماذا لا؟

٣- مــا هي الصناعــات الأخرى التي يمكنها اســتخدام مقاييس تشــغيلية ولوحات معلومات مشابهة؟

Sources: Customers.microsoft.com. (2015). Rockwell Automation: Fueling the oil and gas industry with IOT. https://customers.microsoft.com/Pages/CustomerStory.aspx?recid=19922 (accessed August 2016); Microsoft.com. (n.d.). Customer stories| Rockwell Automation. https://www.microsoft.com/en-us/cloud-platform/cus- tomer-stories-rockwell-automation (accessed July 2016).

البنية التحتية لتقنية إنترنت الأشياء:

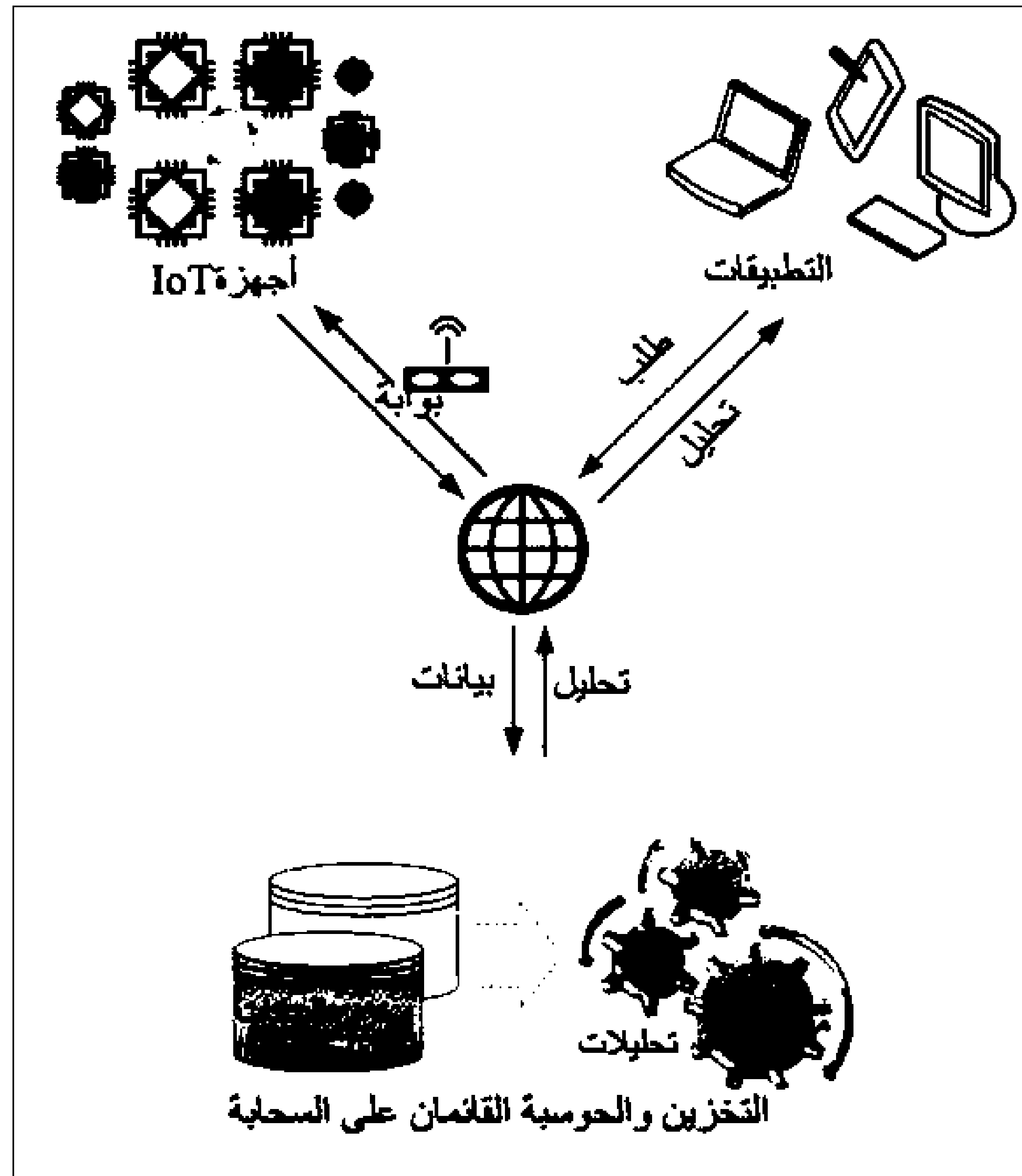
من نظرة بعين الطائر، يمكن تقسيم تقنية إنترنت الأشياء إلى أربع مجموعات رئيسة والموضحة في الشكل (٨-١) كما يلي:

١- الأجهزة (Hardware): وتشــمل الأجهزة المادية، وأجهزة الاستشــعار، والمشغلات الميكانيكة التي يتم فيها إنتاج وتسجيل البيانات. والأجهزة هي المعدات التي تحتاج إلى التحكم فيها أو مراقبتها أو تتبعها. وقد تحتوي أجهزة الاستشــعار الخاصة بتقنية إنترنت الأشــياء على معالج أو أي جهاز حوسبة والذي يقوم بتوزيع البيانات الواردة.

٢- الاتصال (Connectivity): ينبغي أن يكون هناك محطة أساسـية أو مركز رئيس يقوم بجمع البيانات من القطع المزودة بأجهزة الاستشـعار وإرسـال تلك البيانات إلى السحابة. يتم ربط الأجهزة بشـبكة لتتواصل مـع بعضها البعض أو مع تطبيقات أخـرى، وقد يكون هذا الربط مباشر أو غير مباشر بالإنترنت. وتعمل البوابة على تمكين الأجهزة غير المتصلة مباشرةً بالإنترنت من الوصول إلى منصة السحابة.

٣- الخلفيـة البرمجيـة (Software backend): وفي هـذه الطبقة، تتـم إدارة البيانات التي تم جمعها؛ حيث تقوم الخلفية البرمجية بإدارة الشـبكات والأجهزة وتوفير تكامل البيانات. وقد يكون هذا جيدًا جدًا في السحابة (انظر القسم التالي ثانيةً).

٤- التطبيقات: وفي هذا الجزء من إنترنت الأشياء، يتم تحويل البيانات إلى معلومات ذات معنى. وقد يتم تشغيل العديد من التطبيقات على الهواتف الذكية وأجهزة التابلت وأجهزة الحاسب الشـخصي وتقوم بعمل شي مفيد مع البيانات. كما يتم تشـغيل تطبيقات أخرى على الخادم ومن ثم تقديم نتائج أو تنبيهات من خلال لوحات المعلومات أو الرسائل إلى المختصين.



شكل ٨-١: بناء كتل البنية التحتية لتكنولوجيا IoT

أجهزة استشعار التعريف بترددات الراديو RFID:

تعتبر تقنية التعريف بترددات الراديو(RFID) واحدة من أقدم تقنيات الاستشــعار الموجودة والتي تشــهد نموًا كبيرًا. وتعتبر RFID تقنية شــاملة تشــير إلى اســتخدام موجات تردد الراديو لتحديد الأهداف. وبشــكل أســاسي، تعد هذه التقنية مثالًا لعائلة تقنيات تحديد الهوية تلقائيًا، والتي تشــمل أيضًا الباركود والقضبان المغناطيســية واسعة الانتشــار، حيث استخدمت سلاسل توريد التجزئة (والعديد من المجالات الأخرى) منذ منتصف الســبعينيات تقنية الباركود كشــكل أســاسي لتحديد الهوية تلقائيًا. وقد روجت مزايا RFID المحتملة لانتهاج هذه التقنية بقوة إلى العديــد من الشركات (بقيادة كــبرى شركات التجزئة مثــل Wal-Mart، وTarget، وDillard's) كطريقة لتحسين سلاسل التوريد الخاصة بهم وبالتالي تقليل التكاليف وزيادة المبيعات.

كيف تعمل تقنية RFID؟ في أبسط صورة له، يتكون نظام RFID من علامة (متصلة بالمنتج المراد التعرف عليه)، وأداة اســتجواب (بمعنى قارئ)، وهوائي واحد أو أكثر متصلين بالقارئ، وجهاز حاسب (للتحكــم في القــارئ والتقاط البيانات). وحاليًا، بدأت سلاســل التوريد في الاهتمام أكثر باســتخدام علامات RFID غير المباشرة؛ إذ تســتقبل العلامات غير المبــاشرة الطاقة من المجال الكهرومغناطيسي الذي أنشــأته أداة الاســتجواب (وليكن القارئ) والمعلومات المرتدة فقط عندمــا يتم طلبها. وتظل العلامة غير المباشرة نشطة فقط عندما تكون داخل المجال المغناطيسي لأداة الاستجواب.

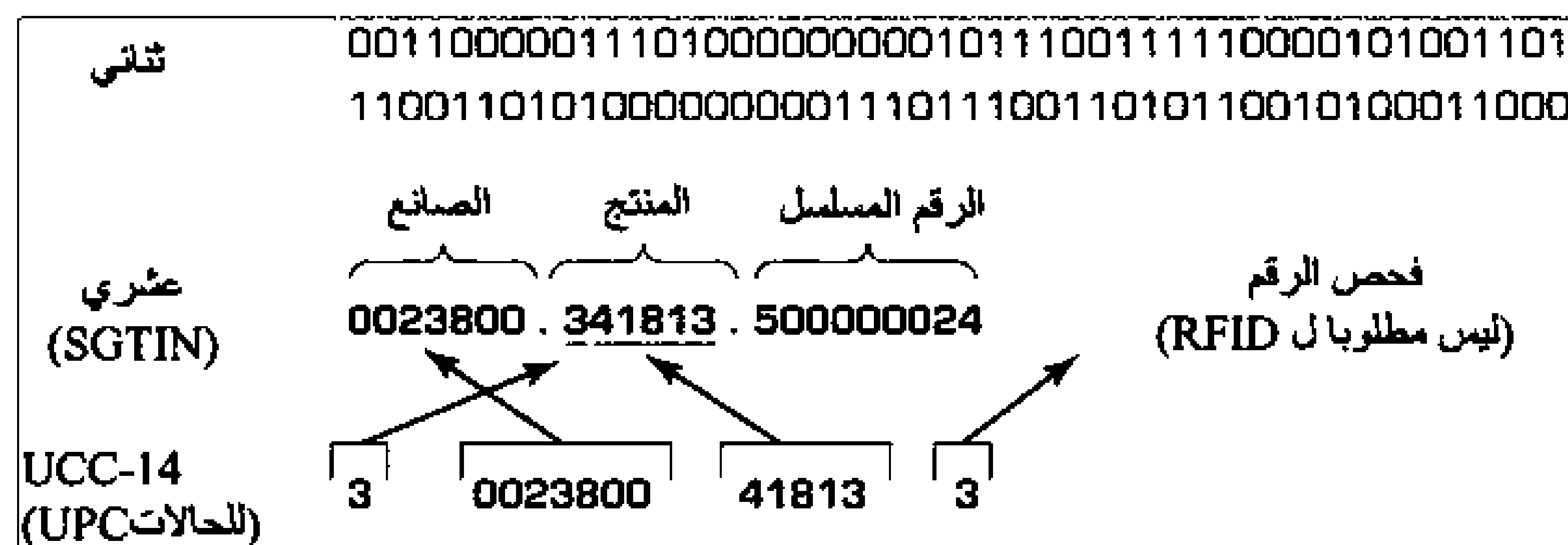
وفي المقابل، يكون لدى العلامات المباشرة بطارية على لوح مسطح لتنشيطها. وحيث إن لدى العلامات المباشرة مصدر طاقة خاصًا بها، فهي لا تحتاج قارئًا لتنشيطها؛ بل يمكنها الشروع في عملية نقل البيانات من تلقاء نفســها. وبالمقارنة مع العلامات غير المباشرة، فإن للعلامات المباشرة نطاق قراءة أطول ودقة أفضل وتخزين معلومات أكثر تعقيدًا قابلًا لإعادة الكتابة، وقدرات معالجة أكثر ثــراءً (Moradpour & Bhuptani, 2005). ولكن نظرًا لوجود البطارية، يكون للعلامات المباشرة بعــض الســلبيات، ومنها أن فترة حياتها تكون أقصر، وحجمها أكــبر، وثمنها أغلى من العلامات غير المباشرة. وحاليًا، يتم تصميم وتشــغيل معظم تطبيقات بيع التجزئة بعلامات ســلبية. وغالبًا ما يتكرر وجود العلامات المباشرة في أنظمة الدفاع أو العســكرية، ولكنها تظهر كذلك في تقنيات مثل EZ Pass، حيث ترتبط العلامات بحساب مدفوع مقدمًا، مما يُمكِّن السائقين من دفع الرسوم أثناء القيادة فقط بالمرور على القارئ بدلًا من التوقف للسداد عند كشك تحصيل الرسوم.

يعد تمثيل البيانات الأكثر انتشــارًا بالنســبة لتقنية RFID هو رمز المنتج الإلكتروني (EPC)، والــذي ينظــر إليه العديد من العاملــين في الصناعة على أنه الجيل التالي مــن رمز المنتج العالمي

(UPC)، والــذي يتم تمثيله في الغالب بما يســمى باركود. وكما هوالحال في UPC، يتكون EPC من سلسلة من الأرقام التي تحدد أنواع المنتجات والمصنعين عبر سلاسل التوريد. كما يتضمن رمز EPC مجموعة إضافية من الأرقام لتحديد العناصر بشكل فريد.

وفي الوقــت الحــالي، تحتوي معظم علامات RFID على ٩٦ بايت من البيانات في شــكل أرقام تعريف التجارة العالمية المسلســلة (SGTINs) لتحديد حالات أو رموز حاوية الشــحن المسلسلة لتحديد لوحات التنزيل (على الرغم من إمكانية استخدام SGTINs أيضًا لتحديد لوحات التنزيل). ويمكــن إيجــاد الدليــل الكامل لوضع علامــة معايير البيانات عــلى موقع الويــب الخاص بشركة EPCglobal وهــو(epcglobalinc.org). وتعتــبر شركة EPCglobal منظمة يشــترك في قيادتها رواد الصناعة والمنظمات التي تركز على إنشاء معايير عالمية لـ EPC لدعم استخدام RFID.

وكما هو موضح في الشــكل (٨-٢)، تظهر بيانات العلامات في أنقى صورها عبارة عن سلســلة من الأرقــام الثنائيــة، والتي يمكــن تحويلها بعد ذلك إلى المعــادل العشري SGTIN. وكــما يتضح، يعتبر SGTIN بشــكل أســاسي UPC (UCC-14 لتحديد حاوية الشــحن) برقم مسلســل. ويعتبر الرقم المسلســل هو أهم فرق بين كل من الـ UPC المكون من ١٤ رقمًا المستخدم اليوم وبين SGTIN الذي يتضمن علامة RFID. فمع UPCs، يمكن للشركات تحديد عائلة المنتج التي تنتمي إليها الحالة (مثل، مناديل كارمن ٨ - عبوات)، ولكن لا يمكن تمييز عبوة واحدة من أخرى. أما مع SGTIN، فيتم تحديد كل قطعة بشكل فريد، وهذا يوفر مستوى رؤية على مستوى القطعة، بدلًا من مستوى عائلة المنتج.



شكل ٨-٢: علامة بيانات RFID

ويعد أحد تطبيقات الكميات الضخمة من البيانات المتولدة عن طريق RFID في إدارة سلاســل التوريد (Delen, Hardgrave, & Sharda, 2007). كما تستطيع الشركات استخدام RFID لتحسين كفاءة أو فعالية العمليات الحالية المتنوعة عن طريق التغيير التدريجي للعملية. فعلى سبيل المثال، اقترحت الأدلة الأولية إمكانية تخفيض وقت اســتلام المنتج في المســتودع باســتخدام (Katz, 2006)

RFID. فبدلًا من عمل مســح ضوئي لكل قطعة من المنتج على انفراد باســتخدام ماســح الباركود، يمكن قراءة المنتج ذي علامة RFID تلقائيًا على بوابة استلام المنتجات. وقدمت شركة جيليت تقريرًا بانخفاض وقت اســتلام لوحات التنزيل في مركز التوزيع الخــاص بها من ٢٠ ثانية إلى ٥ ثوانٍ بفضل RFID وإســتراتيجية علامــة - على - المصــدر التابعة لها (Katz, 2006). ولم تتغير عملية الاســتلام جذريًا (أي إن الرافعات الشــوكية قامت بتفريغ المنتج كما كان من قبل). وكان التغيير الوحيد هو إزالة الحاجة للمسح اليدوي للمنتج. وبالتالي، أصبحت العملية أكثر كفاءة. كما يمكن زيادة فعالية العمليــات. فمثــلًا، وجدت شركة Wal-Mart انخفاضًا بنســبة ٢١٪ في المخزون باســتخدام بيانات RFID لتوليــد قوائم أفضل للمنتجات التي ســيتم تجديدها(Hardgrave, Langford, Waller, & Miller, 2008)؛ إذ لم تتغير عملية تجديد الرف، ولكن تم تحسينها باستخدام RFID. كما يستخدم RFID لتقليل عدد الأخطاء، مما يحسن دقة المخزون، ويؤدي في النهاية إلى تحسين التنبؤ والتجديد.

وقد تم استخدام بيانات RFID في العديد من التطبيقات المرتبطة الأخرى. فعلى سبيل المثال، تمثل البضائع القابلة للتلف أحد التحديات الكبرى لإدارة سلسلة التوريد بسبب العدد الكبير من المتغــيرات ذات الخصائص المختلفــة القابلة للتلف، ومتطلبات حســاب تدفق البضائع في بعض سلاســل التوريد، وكميات كبيرة من البضائع التي يتم تداولها على مســافات طويلة. وعلى الرغم مــن أن الطعام يمثل الجزء الرئيس من قائمة المنتجات القابلة للتلف، إلا أن العديد من المنتجات الأخرى، كالزهور الحديثة والأدوية ومستحضرات التجميل وقطع غيار السيارات، وغيرها، تتطلب ضوابــط بيئيــة صارمة للإبقاء على جودتها. ونظرًا لأن حجم البضائــع المتداولة كبير جدًا، فتزداد احتمالية حدوث المشاكل (Sahin, Babaï, & Dallery Vaillant, 2007). كما يزيد القضاء على نسبة صغيرة من التلف مثلًا، من معنوية التحسن في سلاسل التوريد. ولذلك' تكون للإدارة المثلى لسلاسل توريد البضائع القابلة للتلف أهمية كبرى للشركات في هذا القطاع السوقي.

يعتمد نجاح سلاسل التوريد سريعة الزوال اليوم على مستوى (وتوقيت) بروز المنتج. وينبغي أن يجيــب هذا الظهور على أســئلة «أين المنتج الخاص بي؟» و«مــا هو شرط المنتج الخاص بي؟». وبالفعل، شرعت العديد من الشركات في تجربة استخدام RFID للمواد القابلة للتلف، مع الوضع في الاعتبار كلًا من الأمثلة التالية:

- قامت شركة ساموورث برذارز (Samworth Brothers) للتوزيع (المملكة المتحدة؛ سندويشات وحلويات، إلخ) بإجراء مراقبة فورية لدرجة الحرارة في شاحناتها (Swedberg, 2006a).
- تقوم ســتاربكس باســتخدام تتبع درجات الحرارة لمنتجات إعداد الطعام المتجهة إلى منافذ بيع التجزئة (Swedberg, 2006b).

- تقوم Sysco باستخدام RFID لفحص شروط التنزيل دون فتح الأبواب (Collins, 2005).

ولنا في إدارة جودة المنتج مثال آخر على اســتخدام RFID في سلاســل التوريد. فقد كشفت الدراســات التي تســتخدم علامات RFID القائمة على أجهزة الاستشــعار في الشــاحنات المزودة بمــبردات والمخصصة لحمل المــواد الغذائية أن درجة الحرارة فيهــا لم تكن ثابتة كما هو مفترض، بل تفاوتت بشكل كبير (Delen, Hardgrave, & Sharda, 2011). فقد تتغير البيئة مع انتقال المنتج عبر سلسلة التوريد، وهذا يؤثر على جودة وسلامة المنتج. وهنا يأتي دور أجهزة الاستشعار البيئية التي تدعم 'RFID إذ توفر رؤية للظروف البيئية المتغيرة المحيطة بالمنتج كما توفر البيانات اللازمة لتحديد مدى تأثير تلك التغييرات على جودة أو ســلامة المنتج. وبدون أجهزة استشــعار، يمكننا الحصول على تقديرات متنوعة للظروف البيئية ولكن عند نقطة واحدة فقط دون أي رؤية لما بين تلك النقاط (على ســبيل المثال، درجة الحرارة في وقت التنزيل، وفي وقت التســليم). وفي تطبيقات العينة، تختلف درجات الحرارة وفقًا لأشياء معينة، منها:

- الموقع على لوح التنزيل (في الأعلى، في المنتصف، في الأسفل).

- ترتيب التنزيل (أي موقع لوحات التنزيل).

- نوع الحاوية.

- نوع المنتج.

- مواد التغليف (صندوق مموج أو بلاستيك مقوى).

ويقترح الأثر الملحوظ للعديد من المتغيرات ضرورة مراقبة البيئة بشــكل مســتمر لتحقيق الفهم التام للظروف المحيطة على مســتوى لوح التنزيل و / أو القطعة. وبشــكل عام، فإن أجهزة استشعار (درجــة الحــرارة) الممكّنة بتقنيــة RFID تعمل جيدًا وتوفر رؤية واســعة لا يمكــن الحصول عليها بتقديــرات النقطــة الواحدة للظروف التي تواجــه المنتج عند مروره عبر سلاســل التوريد. والدرس المستفاد بشكل عام هو أن تقنية RFID تولد كميات ضخمة من البيانات التي يمكن تحليلها للحصول على أفكار عظيمة بشأن بيئة الشركة، وهذا هو الغرض الرئيس لقيام ذكاء الأعمال ودعم القرار.

الحوسبة الضبابية:

تعد إحدى أهم القضايا في إنترنت الأشياء هي أن البيانات التي تنتجها أجهزة الاستشعار بيانات ضخمة، ولكن ليســت كلهــا مفيدة. إذًا، ما هــو القدر الواجب رفعه مــن البيانات الضخمة على

الخوادم السحابية لتحليلها؟ وتعتبر فكرة الحوسبة الضبابية هي المفهوم الجديد لحل هذا السؤال؛ حيث يقوم الضباب بتمديد السحابة لتكون أقرب إلى الأشياء التي تُنتِج وتعمل على بيانات إنترنت الأشياء. ويمكن وضع هذه الأجهزة، والتي يطلق عليها عقد الضباب، في أي مكان بين ربط الشبكة. ويمكن لأي جهاز يتوفر له الحوسبة والتخزين والاتصال بالشبكة أن يكون عقدة ضباب، كأجهزة الراوتر أو المحولات. وفيما يلي توضيح لذلك:

مركز البيانات / السحابة ← جهاز الضباب ← جهاز مادي / أجهزة استشعار لتوليد البيانات

يؤدي تحليل البيانات بالقرب من الأجهزة إلى تقليل وقت الاستجابة. كما يحافظ أيضًا على السعة التنزيلية، إذ إن إرسال البيانات إلى السحابة يتطلب سعة تحميلية كبيرة. وتزداد أهمية الحوسبة الضبابية عندما تحتاج البيانات للتحليل في أقل من ثانية، كما هو الحال في حالة عطل النظام المتتالي. ويعرض الجدول (٨-١) فرقين بسيطين بين السحابة والضباب.

جدول ٨-١: الفرق بين عقد الضباب ومنصة السحاب

عقد الضباب	منصة السحابة
تستقبل البيانات من أجهزة إنترنت الأشياء.	تستقبل وتجمع البيانات من عقد الضباب.
تقوم بتنفيذ تحليلات إنترنت الأشياء في وقت الاستجابة بالمللي ثانية.	يتم إجراء التحليل على كميات ضخمة من بيانات الأعمال والذي قد يستغرق ساعات أو أسابيع.

كما قد تعطي الحوسبة الضبابية مزيدًا من الأمان، حيث يمكن تأمين عقد الضباب باستخدام نفس حل الأمان المستخدم في بيئات تقنية المعلومات الأخرى.

منصات إنترنت الأشياء:

ولأن إنترنت الأشياء لا يزال قيد التطوير، فإن العديد من منصات التقنية الخاصة بالمجال والتطبيقات تتطور أيضًا. وليست مفاجأة أن يكون العديد من الموردين الأساسيين لأنظمة إنترنت الأشياء هم نفس الموردين الذين يوفرون خدمات تحليل ومستودعات البيانات لمجالات التطبيق الأخرى، والتي منها:

- Amazon AWS IoT.
- Microsoft Azure IoT Suite.
- Predix IoT Platform by General Electric (GE).
- حلول IBM Watson IoT.
- Teradata Unified Data Architecture.

وقد عرضنا مثالًا على ذلك في المقالة الافتتاحية القسم ٨-١. كما تقدم حالة عملية ٨-٣ تطبيقًا من منصة رئيسة أخرى لإنترنت الأشياء والتي تقوم بتسويقها جنرال إلكتريك.

حالة عملية ٨-٣

التعاون المشترك بين بيتني باوز Pitney Bowes ومنصة جنرال إلكتريك General Electric IoT لتحسين الإنتاج

تعتبر Pitney Bowes، ومقرها الرئيس في ولاية كونيتيكت Connecticut، هي المورد الأمريكي لحلول التجارة الإلكترونية ومنتجات الشحن والبريد. تقوم آلات الشركة بتشغيل أو إنتاج بمعدل ١٥٠ مليون قطعة بريد في اليوم. ويقوم Enterprise Mail Business التابع لـ Pitney Bowes بتطوير وبيع وخدمة آلات كبيرة تسمى Inserters. تساعد هذه الآلات البنوك ومقدمي الرعاية الصحية على وضع قطعة البريد الصحيحة في المظروف ذي المقاس الصحيح. بالنسبة لبنك واحد، قامت هذه الآلة بتجميع ٩٠٠ مليون قطعة بريد في السنة. وتنتج معظم الآلات ٢٠٠٠٠ رسالة في الساعة. وبهذا المستوى من الإنتاج الآلي، تقوم Pitney Bowes وعملاؤها أيضًا بإنتاج كمية ضخمة من البيانات كل يوم. وتسعى Pitney Bowes لتحليل البيانات المتولدة من آلات البريد هذه للتنبؤ بأعطال الآلات مقدمًا ومن ثم منع أي انقطاع للخدمة وإصلاح الآلات قبل أن تتعطل. وقد اعتمدت GE Predix Pitney Bowes، كمنصة سحابة مصممة للإنترنت الصناعي؛ إذ تقوم منصة Predix بتحليل بيانات جهاز الاستشعار للآلات وباستخدام التحليلات الفورية، تقدم أفكارًا قوية تسهل اتخاذ القرارات. وبالإضافة إلى منع انقطاع الخدمة، تساعد التحليلات المقدمة من Predix على تقليل وقت العطل وبالتالي زيادة الإنتاجية. وقد أدى التحليل الشامل لبيانات جهاز الاستشعار الذي تجريه Predix إلى زيادة الإنتاجية وتحسين أداء الآلات.

Sources: News.pb.com. (2015). GE and Pitney Bowes join forces to bring the power of the industrial Internet to the world of commerce. http://news.pb.com/article_display.cfm?article_id=5634 (accessed August 2016); GEreports.com. (2016). The power of Predix: An inside look at how Pitney Bowes is using the industrial Internet platform. http://www.gereports.com/the-power-of-predixan-inside-look-at-how-pitney-bowes-has-been-using-the-industrialinternet-platform/; (accessed August 2016); GE.com (2016). Making machines intelligent is smart business. http://www.ge.com/digital/sites/default/files/ge_digital_predix_pb_brochure.pdf; (accessed August 2016); GE.com. (2015). The

industrial Internet, Pitney Bowes and GE. http://blogs.pb.com/corporate/201514/07//the industrialinternet-pitney-bowes-and-ge/(accessed August 2016).

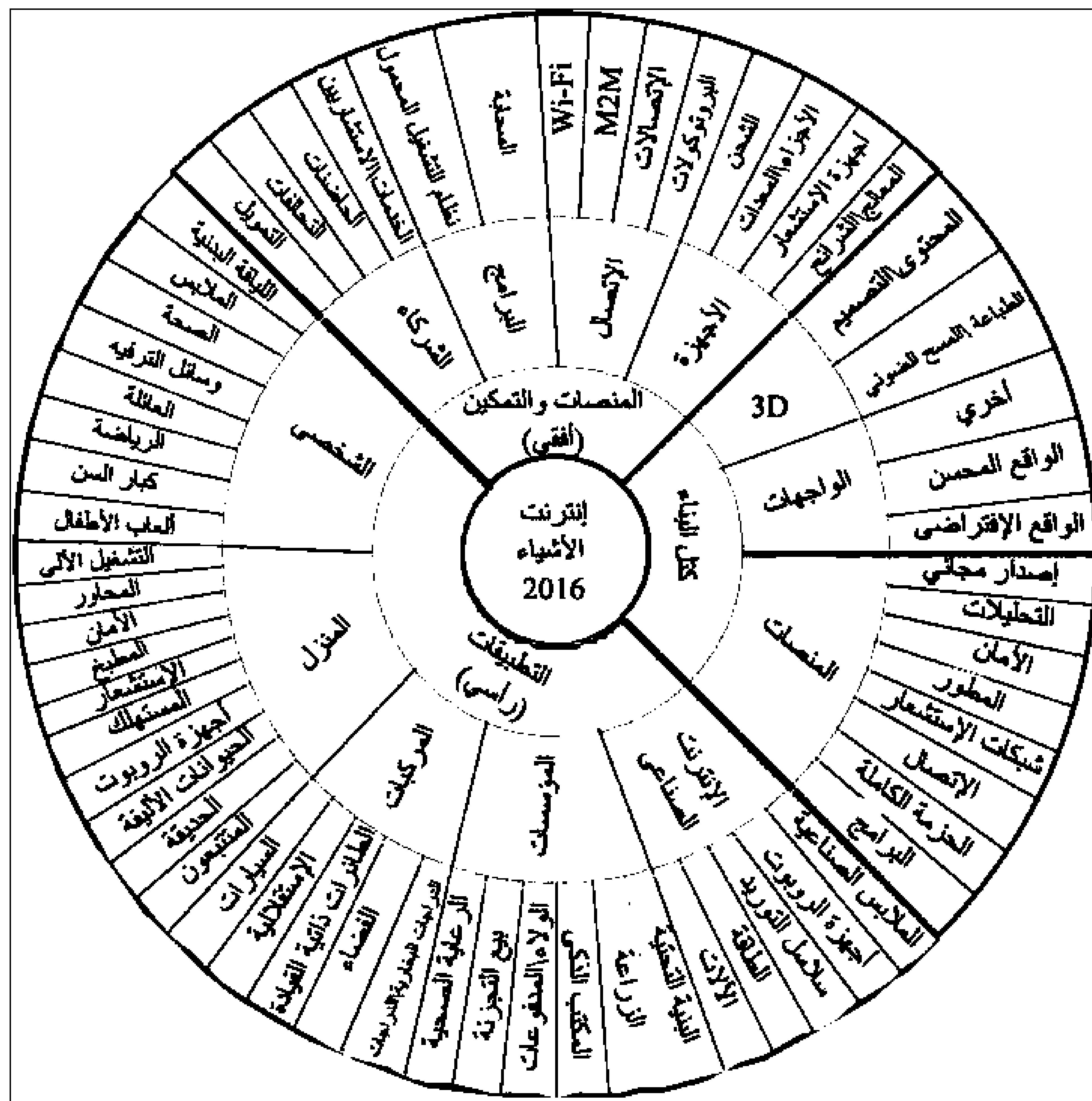
بدء النظام البيئي لإنترنت الأشياء:

تشرع العديد من الشركات المبتدئة في الظهور في مجال إنترنت الأشـياء، وفي غضون ثلاث إلى أربع سـنوات قادمة سـنرى ازدهارًا كبيرًا في صناعة إنترنت الأشياء. وقد نما رأس المال المطروح في مشـهد إنترنت الأشياء، من ١,٨ بليون دولار في عام ٢٠١٣ إلى ٢,٥٩ بليون دولار في عام ٢٠١٤، ثم إلى ٣,٤٤ بليون دولار في عام ٢٠١٥. ولعل من أبرز الأمثلة على الشركات الناشئة في إنترنت الأشياء: Sigfox و3D Robotics وCanary وAthos وGreenwave وJawbone وFreedomPop و Razer وRing.

وتعـد Fitbit من أنجح شركات إنترنت الأشـياء، وهي شركة أمريكيـة تقوم بتصنيع متتبعات النشـاط، وأجهزة تقنية لاسلكية تستخدم أجهزة استشـعار لتحديد عدد خطوات المشي، ومعدل القلب، وجودة النوم، وخطوات التسلق، وغيرها من مقاييس الصحة الشخصية.

كما تقوم شركات الاتصالات أيضًا باستكشـاف وترويج إنترنت الأشـياء. ويسعى معظم كبار المشـتغلين بالاتصالات مثل AT&T إلى الاسـتفادة من هذه التقنية. ولتوفير الاتصال بالسيارات، أبرمت AT&T شراكة مع ٨ شركات أمريكية لتصنيع السيارات من أصل ١٠ شركات تعمل في هذا المجال. وتعرض العديد من شركات الاتصالات شبكات الجيل الخامس القادمة الخاصة بها كحجر أساس لإنترنت الأشياء.

يعتبر كل من Google/Alphabet وAmazon هما الأبرز في النظام البيئي لإنترنت الأشـياء. كما تعد Google بسحابة Google التابعة لها، وSideWalk Labs (مدن ذكية) والسيارات ذاتية القيادة مسـتثمرًا رائدًا في العديد من مبادرات إنترنت الأشياء. وقد أنتجت Amazon باستخدام خدمات الويب أمازون (AWS) التابعة لها منصة جديدة لإنترنت الأشـياء والتي يمكن أن تعمل كخلفية لإنترنت الأشـياء. ويقدم الشـكل (٨-٣) عرضًـا موجزًا لمكونات النظـام البيئي لإنترنت الأشياء، والذي يشتمل على مجالات التطبيق المختلفة، وصانعي الأجهزة، ومقدمي خدمة الاتصال، ومطوري البرمجيات، ومستشاري التحليلات، وما إلى ذلك، وهو مقتبس من مخطط النظام البيئي الذي تـم تطويره وصيانته عـلى المدونـة: -http://mattturck.com/2016/03/28/2016-iot landscape/ من قِبَل صاحبها Matt Turck (تمت الزيارة في أغسطس ٢٠١٦).



Sources: Adapted from matt,T(March, 2016). Internet of Things: Are We There Yet? http://mattturck.com/2016/03/28/2016-iot-landscape/ (accessed August 2016) .

شكل ٨-٣: النظام البيئي لإنترنت الأشياء

يتضمن هذا المخطط المفصل قائمة لشركات في كل مجموعة. ويبين الشـكل (٨-٣) أن فرص إنترنت الأشـياء تتطور في ثلاثة قطاعات رئيسة؛ إذ تشتمل المجموعة الأولى على الشركات التي توفر مجموعات بناء تقنيات وأدوات تمكين إنترنت الأشـياء (الأجهزة والبرمجيات والاتصال والمستشـارين / الحاضنات / التحالفات / الشركاء). وتشـمل المجموعة الثانيـة المنصات والواجهات والطباعة ثلاثية الأبعاد الحديثة وموردي التقنية / الخدمة المتحدين، ويمكننا أن نطلق على هذه المجموعة اسم (موردو الخدمة لمطوري إنترنت الأشياء - قطاع أفقي)، في حين تتألف المجموعة الثالثة من قطاعات الصناعة، وهي المنطقة التي ستكون إلى حد كبير تطبيقات في قطاعات مختلفة، وهي القطاع الشخصي والمنزل والمركبات والمؤسسات والإنترنت الصناعي، حيث يقدم الشـكل (٨-٣) صورة واحدة لهذه المنظمة، بينما يتضمن موقع المدونة صورة مختلفة تتضمن أمثلة لشركات في كل من المجموعات الفرعية المحددة هنا.

اعتبارات إدارية في إنترنت الأشياء:

على الرغم من الإثارة الضخمة الدائرة حول نمو وإمكانية إنترنت الأشـياء، إلا أن هناك بعض القلـق الـذي ينبغي على المديرين الإلمام بـه. وقد جمع معهد ماكنـزي (McKinsey) العالمي (٢٠١٥) دليلًا تنفيذيًّا ممتازًا لإنترنت الأشياء.

ويحدد هذا التقرير القضايا التالية:

١- التوافق التنظيمي: على الرغم من صحة ذلك بالنسـبة للكثير من مبادرات التقنية الأخرى، فمع إنترنت الأشياء، فإن فرص التحسينات التشغيلية وإيجاد فرص أعمال جديدة تعني أنه يجب على موظفي تقنية المعلومات وموظفي التشغيل أن يعملوا كفريق واحد وليس في وظائف منفصلة. وكما لاحظ مؤلفو الدليل، «فإن إنترنت الأشـياء سـوف يتحدى المفاهيم الأخرى للمسـؤوليات التنظيمية. كما ينبغي على مديري أقسـام المالية والتسـويق والتشـغيل، بجانب قادة وحدات الأعمال، أن يكونوا متقبلين لربط أنظمتهم» (McKinsey, 2015).

٢- تحديات العمل المشـترك: يشكل العمل المشترك ضررًا كبيرًا حتى الآن في نمو هذه الأنظمة؛ إذ لا تتصل جميع الأجهزة مع بعضها البعض بسلاسـة. ثانيًا، هناك الكثير من المشـاكل التقنية في الاتصـال، ولا تملك العديد من المناطق النائية اتصالات Wi-Fi مناسـبة بعد، وأيضًا تكون القضايـا المتعلقة بمعالجة البيانات الضخمة مسـؤولة عن التقدم البطيء في إنترنت الأشـياء. وتحاول الشركات تقليل البيانات على مسـتوى جهاز الاستشـعار بحيث يتم تمرير الحد الأدنى فقط من البيانات إلى السـحابة، مما يجعل دعم البنيـة التحتية الحالية للكمية الضخمة من البيانات المتولدة من إنترنت الأشـياء أمرًا شـاقًّا. وهناك مشكلة ذات صلة أخرى وهي تزويد الأجهـزة بأجهزة استشـعار لتكون قادرة على جمع ونقل البيانات للتحليل، ويسـتغرق الأمر في المتوسـط من ١٨ إلى ٢٤ شهرًا لتبدأ إنترنت الأشـياء بشحن منتجها، ثم من سنة إلى سنتين إضافيتين لتوزيع وبيع منتجها في السـوق. وبالإضافة إلى ذلك، سيسـتغرق المستهلكون وقتًا لاسـتبدال الأشـياء المتناظرة بمنتجات ذكية جديدة من إنترنت الأشياء. وكمثال، فإنه بالنسبة للناس يعد اسـتبدال الهواتف الجوالة أسـهل من استبدال السـيارة وأدوات المطبخ والأقفال وأشياء أخرى يمكن أن تستفيد من وجود جهاز استشعار وبقائه متصلًا مع إنترنت الأشياء.

Source: Cisco.com. (2015). Fog computing and the Internet of Things: Extend the Cloud to where the things are. https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf (accessed August 2016).

٣- الأمــان: يمثل أمن البيانات قضية عامة، ولكنها في ســياق إنترنت الأشــياء تعتــبر قضية أكبر؛ إذ يصبح كل جهاز متصل بـ إنترنت الأشــياء نقطة دخول أخرى للقراصنة الأشرار للوصول إلى نظام كبير، أو على الأقل تشغيل أو إتلاف الجهاز المحدد. فهناك قصص لقراصنة قادرين على الاختراق والتحكم في الوظائف الآلية للســيارة، أو السيطرة على فتح باب الجراج عن بعد. وتتطلب مثل هذه القضايا أن ينطوي أي اعتماد واسع النطاق لإنترنت الأشياء على اعتبارات أمنية منذ البداية.

وعلى الرغم من هذه الاعتبارات الإدارية، إلا أن النمو الناشــئ في إنترنت الأشــياء وإمكانياته لمســاعدتنا في تحقيق الرؤية للمدن الذكية وللشــبكة الذكية ولأي شيء ذكي يعد أمرًا محيرًا. فهي إحدى المجالات المفتوحة الشاسعة للإبداع وريادة الأعمال.

أسئلة مراجعة على القسم ٨-٢:

١- ما هي استخدامات إنترنت الأشياء الأساسية؟

٢- ما هي مجموعات البناء التقنية في إنترنت الأشياء؟

٣- ما هي RFID؟

٤- إبحث على الإنترنت عن تطبيقات RFID في الرعاية الصحية والترفيه والرياضة.

٥- حدد بعض أهم العاملين في النظام البيئي لإنترنت الأشياء، واستكشف عروضهم.

٦- اذكر بعض القضايا الرئيسة التي ينبغي على المديرين إدراكها عند استكشاف إنترنت الأشياء؟

٨-٣ الحوسبة السحابية وتحليلات الأعمال:

يوجــد اتجاه تقني حديث الظهور ينبغي على مســتخدمي تحليــلات الأعمال الإلمام به وهو الحوســبة الســحابية. ويعرِّف المعهد الوطني للمعايير والتقنية (NIST) الحوسبة السحابية على أنها: «نموذج لتمكين الوصول إلى الشــبكة المناســبة عند الطلب إلى مجموعة مشتركة من موارد الحوســبة القابلة للتشــكيل (كالشــبكات والخوادم والتخزين والخدمات) والتــي يمكن توفيرها وإصدارها بسرعة وبأقل جهد إداري أو تفاعل من مورد الخدمة»، في حين تُعرِّف ويكيبيديا (n.d., Cloud Computing) الحوســبة الســحابية تعريفًا فضفاضًا وشاملًا وهو: "أسلوب حوسبة يتم من خلاله توفير الموارد القابلة للتطوير بشــكل ديناميــكي وغالبًا ما تكون افتراضية ويتم توفيرها عبر الإنترنت، ولا يحتاج المســتخدمون إلى معرفة أو خــبرة في أو التحكم في البنى التحتية للتقنية في الســحابة التي تدعمهم". ومن ناحية أخرى، تعد الحوســبة الســحابية اســمًا جديدًا للعديد

من الاتجاهات السـابقة ذات الصلة، كحوسبة المرافق، وحوسـبة شبكة مقدم خدمات التطبيق، والحوسـبة عند الطلب، والبرمجيات كخدمة (SaaS)، وحتى الحوسبة المركزية القديمة باستخدام المحطات الصامتة. غير أن مصطلح الحوسبة السحابية ينتج أصلًا من إشارة إلى الإنترنت باعتباره «سـحابة» ويمثل تطورًا في جميع اتجاهات الحوسـبة المشتركة / المركزية السابقة. ويعرف القائم بالإدخال في ويكيبيديا أيضًا أن الحوسبة السحابية هي مزيج من عدة مكونات لتقنية المعلومات كخدمات. فعلى سـبيل المثال، تشير البنية الأساسـية كخدمة (IaaS) إلى توفير منصات الحوسبة كخدمـة (PaaS)، بالإضافة إلى توفير منصة أساسـية، مثـل إدارة الحكومة، والأمان، وما إلى ذلك، وتتضمن أيضًا SaaS والتي تشتمل على تطبيقات يتم نقلها من خلال متصفح ويب، في حين تكون البيانات وبرامج التطبيقات موجودة على خادم آخر.

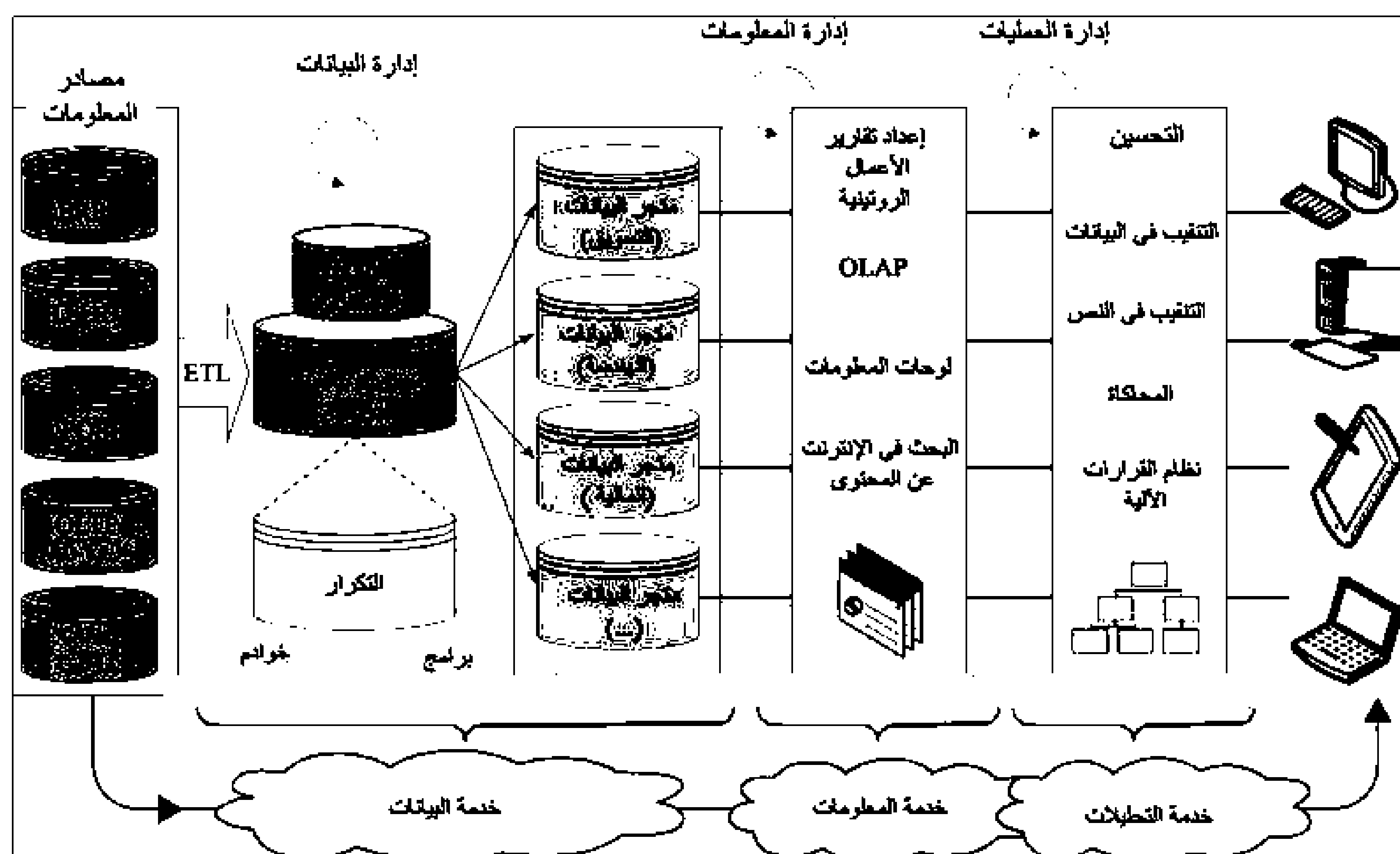
بالرغم من أننا عادةً لا ننظر إلى البريد الإلكتروني على الويب كمثال للحوسبة السحابية، إلا أنه يمكن اعتباره تطبيقًا سحابيًا أساسيًا. وعادةً ما يقوم تطبيق البريد الإلكتروني بمستودعات البيانات (رسـائل البريد الإلكـتروني) والبرمجيات (برامج البريد الإلكتروني التي تسـمح لنا بمعالجة وإدارة رسـائل البريد الإلكتروني). ويوفر مزود البريد الإلكتروني أيضًا الأجهزة / البرامج والبنية الأساسـية بالكامـل. وطالما ظل الإنترنـت متاحًا، يكون الوصول إلى تطبيق البريـد الإلكتروني من أي مكان في السـحابة ممكنًا. وعندما يتم تحديث التطبيق من قبل مزود البريد الإلكتروني (فمثلًا، عندما يقـوم Gmail بتحديث تطبيق البريـد الإلكتروني الخاص به)، فإنه يصبـح متاحًا لجميع العملاء دون حاجتهم لتنزيل أي برامج جديدة. ومن أمثلة الحوسـبة السـحابية أيضًا مواقع الشـبكات الاجتماعية على الويـب مثل Facebook وTwitter وLinkedIn. وبالتالي، يعتبر أي تطبيق عام على الويب مثالًا على تطبيق سـحابي. وهناك مثال آخر على تطبيق السـحابة العامة وهو وثائق وجداول البيانات من Google؛ حيث يسـمح هذا التطبيق للمسـتخدم بإنشـاء وثائق نصية أو جـداول بيانـات والتي يتم تخزينها على خـوادم Google، وتكون متاحة للمسـتخدمين من أي مكان يمكنهم الوصول إلى الإنترنت فيه. ومرةً أخرى، لا يحتاج المسـتخدم تثبيت أي برامج باسـم «التطبيق في السحابة». كما أن مساحة التخزين موجودة أيضًا «في السحابة».

وهناك مثال جيد على الأعمال العامة للحوسبة السحابية وهو خدمات الويب الخاصة بشركة Amazon.com؛ حيث قامت Amazon.com بتطوير بنية تحتية تقنية مذهلة لكل من التجارة الإلكترونية وذكاء الأعمال وإدارة علاقات العملاء وإدارة سلاسل التوريد، كما قامت بإنشاء مراكز بيانات رئيسـة لإدارة عملياتها الخاصة. ومع ذلك، ومن خلال الخدمات السحابية في Amazon.com، يمكـن للعديد من الشركات الأخرى اسـتخدام نفس هذه التسـهيلات للحصول على مزايا

هــذه التقنيــات دون القيام باســتثمار مماثل. وكما هو الحال في خدمات الحوســبة الســحابية الأخرى، يســتطيع المســتخدم أن يشترك في أي من التسهيلات على أســاس: ادفع - كما - تذهب (Pay as you go). ويعد هذا النموذج من الســماح لشــخص آخر بامتلاك الأجهزة والبرمجيات مع صنع استخدام التسهيلات على أساس الدفع - لكل - استخدامًا بمثابة حجر الزاوية في الحوسبة السحابية. ويوفر عدد من الشركات خدمات الحوسبة السحابية، مثل Salesforce.com و(IBM (Bluemix) و(Microsoft Azure) وGoogle وغيرها.

وكــما هو الحال في العديــد من اتجاهات تقنيــة المعلومات الأخرى، فقد أنتجت الحوســبة السحابية عروضًا جديدةً في ذكاء الأعمال؛ حيث تسمح هذه الخيارات للمنظمة بتطوير مستودع بياناتها والدفع فقط نظير ما تســتخدمه بالفعل. وقد يســتخدم المســتخدم النهائي لخدمة ذكاء الأعمال القائمة على الســحابة منظمة واحدة لتطبيقات التحليل والتي تســتخدم في المقابل شركة أخــرى للمنصة أو للبنية التحتية. وتلخــص المقاطع المتعددة التالية أحدث الاتجاهات في واجهة الحوسبة السحابية وذكاء الأعمال / تحليلات الأعمال (Demirkan & Delen, 2013).

ويوضح الشكل (٨-٤) بنية مفاهيمية لبيئة دعم القرار الموجهة بالخدمات، والتي تعتبر نظام تحليلات قائم على الســحابة. وفي هذا الشــكل يتم دمج الخدمات القائمة على السحابة مع بنية التحليلات العامة المقدمة في الفصول السابقة.



Sources: Based on Demirkan,H.,& Delen,D (April, 2013)Ievergingcapabilitiesof-orienteddecision support systtems: putting analytics and big data in cloud. decision support Systtems.55(1).412-421.

شكل ٨-٤: بنية مفاهيمية لنظام دعم سحابي

وفي حلــول دعم القرار الموجهــة بالخدمات، يمكن الحصول على العناصر التالية كلٌ على حدة أو مجتمعة وتقديمها للمستخدمين كخدمة، وهذه العناصر هي:

١- أنظمة التشغيل.

٢- مستودعات البيانات.

٣- المعالجة التحليلية الفورية.

٤- مكونات المستخدم النهائي.

ويمكــن الحصول على إحدى أو كل هذه الخدمات من خلال الســحابة. ولأن مجال الحوســبة الســحابية يتطــور سريعًا وينمو بإيقاع متســارع، فقــد نتج عن ذلك التباس كبــير في المصطلحات المستخدمة من جانب مختلف الموردين والمستخدمين، فتختلف المسميات من حيث البنية الأساسية والمنصة والبرمجيات والبيانات والمعلومــات والتحليلات كخدمة. ونعرِّف فيما يلي هذه الخدمات، ثم نلخص المنصات التقنية الحالية ونسلط الضوء على تطبيقات لكل منها من خلال حالات عملية.

البيانات كخدمة (DaaS):

يدافــع مفهوم البيانات كخدمة بشــكل أســاسي عــن وجهة النظر التي تقــول: «أين تعيش البيانــات»، حيث لا تهتم المنصــة الفعلية بمكان إقامة البيانات. فقــد تتواجد البيانات في جهاز حاسب محلي أو في خادم في مزرعة خوادم داخل بيئة الحوسبة السحابية. ومع DaaS، يمكن لأي عملية أعمال الوصول إلى البيانات أينما وجدت. بدأت البيانات كخدمة بفكرة أن جودة البيانات تحدث في مكان مركزي، وتطهير وإثراء البيانات وعرضها على أنظمة أو تطبيقات أو مســتخدمين مختلفين، بغض النظر عن مكان وجود تلك البيانات ســواء في المنظمة، أو على أجهزة الحاســب، أو على الشــبكة. والآن، تم استبدال هذا بإدارة البيانات الأساسية وحلول تكامل بيانات العملاء، حيث يمكن أن تتواجد سجلات العملاء (أو المنتجات، أو الأصول، إلخ) في أي مكان، وتكون متاحة كخدمة لأي تطبيق به خدمات تســمح بالوصول إليها. ومن خلال تطبيق مجموعة قياســية من التحويلات على المصادر المتنوعة للبيانات (كأن يتم التأكد من أن الحقول الخاصة بالجنس والتي تحتوي على أنماط رمزية مختلفة [مثل، Mr/Ms، M/F] تتم ترجمتها جميعًا إلى ذكر / أنثى) ثم تمكــين التطبيقات من الوصول إلى البيانات عبر واجهات مفتوحة مثل SQL وXQuery وXML، بحيث يمكن لطالبي الخدمات الوصول إلى البيانات دون الحاجة إلى المورد أو النظام.

وباستخدام DaaS، يستطيع العملاء الانتقال بسرعة وذلك لسهولة الوصول إلى البيانات وحقيقة أنهم ليسو بحاجة إلى معرفة شاملة بالبيانات الأساسية. فإذا طلب العملاء بنية بيانات مختلفة إلى حد ما أو متطلبات خاصة بالموقع، تكون الاستجابة لهم سهلة لأن التغييرات تكون عند أدنى حد لها (رشيقة). ثانيًا، يستطيع مقدمو الخدمة بناء القاعدة بالاستعانة بخبراء البيانات ومصادر خارجية لطبقات التحليل أو التقديم (والتي تتيح للمستخدم واجهات فعالة جدًا من حيث التكلفة كما تجعل طلبات التغيير في طبقة التقديم أكثر جدوى)، ويتم التحكم في الوصول إلى البيانات من خلال خدمات البيانات، ويتجه إلى تحسين جودة البيانات لأن هناك نقطة واحدة للتحديثات.

البرمجيات كخدمة (SaaS):

يتيح هذا النموذج للمستهلكين استخدام التطبيقات والبرامج التي يتم تشغيلها على أجهزة الحاسب البعيدة في البنية التحتية للسحابة. ولا يحتاج المستهلكون أن يهتموا بإدارة البنية التحتية للسحابة؛ فيجب عليهم الدفع مقابل استخدام البرمجيات فقط. وكل ما نحتاج إليه هو متصفح ويب للاتصال بالسحابة. وتقدِّر شركة جارتنر أن إيرادات SaaS في عام ٢٠١٥ حوالي ٣٢ مليار دولار أمريكي وتستخدم في ٧٧٪ من جميع المنظمات. وتعتبر كل من Gmail وPicasa و Flickr أمثلة على SaaS.

المنصة كخدمة (PaaS):

وباستخدام هذا النموذج، تستطيع الشركات نشر البرمجيات والتطبيقات الخاصة بها في السحابة بحيث يمكن استخدامها من قِبَل العملاء. ولا يتعين على الشركات إدارة الموارد المطلوبة لإدارة تطبيقاتها في شبكات تشبه السحابة أو الخوادم أو التخزين أو أنظمة التشغيل، وهذا يقلل من تكلفة صيانة البنية التحتية الأساسية لتشغيل برامجهم كما يوفر أيضًا وقت إعداد هذه البنية التحتية. والآن، يستطيع المستخدمون التركيز فقط على أعمالهم دون التركيز على إدارة البنية التحتية لتشغيل برامجهم. ومن الأمثلة على PaaS: Microsoft Azure وAmazon EC2 وGoogle App Engine.

البنية التحتية كخدمة (IaaS):

في هذا النموذج، يتم توفير موارد البنية التحتية كالشبكات والتخزين والخوادم وموارد الحوسبة الأخرى لشركات العملاء؛ إذ يستطيع العملاء تشغيل تطبيقاتهم والحصول على حقوق إدارية لاستخدام هذه الموارد ولكنهم لا يديرون البنية التحتية الأساسية. ويجب على العملاء الدفع مقابل استخدام البنية التحتية. وخير مثال على ذلك هو خدمات الويب في Amazon.

com. فقد طــورت Amazon.com بنية تحتية تقنية رائعة تتضمن مراكز بيانات، بحيث يمكن للشركات الأخرى اســتخدام خدمات الســحابة التابعة لـــ Amazon.com بنظام الدفع - لكل - استخدام دون الحاجة إلى عمل استثمارات مماثلة.

يجب أن نلاحظ أن هناك الكثير من الالتباس والتداخل في اســتخدام مصطلح الســحابة. فعلى ســبيل المثال، يضيف بعض الموردين أيضًا المعلومات كخدمة (IaaS)، والتي تعد امتدادًا لـ DaaS. ومن الواضح أن IaaS تختلف عن «البنية التحتية كخدمة» التي تم شرحها منذ قليل. ويكون هدفنا هنا هو التعرف فقط على وجود درجات متفاوتة من الخدمات التي يمكن لأي منظمة الاشتراك فيها لإدارة تطبيقات التحليلات. ويســلط الشكل (٨-٥) الضوء على مستوى الاشتراكات في الخدمة التي يستخدمها العميل في كل نوع من الثلاثة للعروض السحابية الرئيسة. ويتضح من الشكل أن SaaS تأتي في المســتوى الأعلى من الخدمة الســحابية التي قد يحصل عليها العميل. فمثلًا، عند استخدام Office 365، تستخدم إحدى المنظمات البرمجيات كخدمة، ويكون العميل هو المسؤول الوحيد عن إحضار البيانات، وتقع العديد من التحليلات كتطبيق خدمة في هذه الفئة أيضًا.

البرمجيات كخدمة SaaS	المنصة كخدمة PaaS	البنية التحتية كخدمة IaaS
التطبيق	التطبيق	التطبيق
البيانات	البيانات	البيانات
وقت التشغيل	وقت التشغيل	وقت التشغيل
الوسطاء	الوسطاء	الوسطاء
نظام التشغيل	نظام التشغيل	نظام التشغيل
الإفتراضية	الإفتراضية	الإفتراضية
الخوادم	الخوادم	الخوادم
التخزين	التخزين	التخزين
الشبكات	الشبكات	الشبكات

تدار بواسطة العميل

تدار بواسطة مورد السحابة

شكل ٨-٥: حزمة التقنية كخدمة لأنواع مختلفة من عروض السحابة

التقنيات الأساسية للحوسبة السحابية:

الافتراضية: وهي إنشــاء نســخة افتراضية لشيء يشبه نظام التشــغيل أو الخادم. ومن الأمثلة البســيطة على الافتراضية هو التقســيم المنطقي لمشغل الأقراص الصلبة لإنشــاء مشغلين منفصلين

للأقراص الصلبة في الحاسب. ويمكن للافتراضية أن توجد في جميع المجالات الثلاثة للحوسبة كما يلي:

- افتراضية الشبكة: أي تقسيم السعة التنزيلية المتاحة إلى قنوات، وهذا يعمل على إخفاء تعقيد الشبكة بتقسيمها إلى أجزاء يمكن إدارتها، ثم بعد ذلك يمكن تخصيص كل سعة تحميلية لخادم أو جهاز معين بشكل فوري.
- افتراضيـــة التخزيـــن: هو تجميع التخزين المـــادي من أجهزة تخزين شـــبكة متعددة إلى جهاز تخزين واحد يمكن إدارته من وحدة التحكم المركزية.
- افتراضية الخادم: وهو إخفاء الخوادم المادية عن مستخدمي الخادم، ولا يتعين على المستخدمين إدارة الخوادم الفعلية أو فهم التفاصيل المعقدة لموارد الخادم.

ويرتبط هذا الاختلاف في مستوى الافتراضية مباشرةً مع الخدمة السحابية المستخدمة.

نماذج نشر السحابة:

يمكن الحصول على الخدمات الســحابية بعدة طرق، مــن بناء بنية تحتية خاصة بالكامل إلى المشاركة مع الآخرين. وتعد النماذج الثلاثة التالية هي الأكثر شيوعًا:

- **ســحابة خاصة:** ويطلق عليها أيضًا ســحابة داخلية أو ســحابة الشركات، وهــي صورة للخدمة السحابية أكثر أمانًا من السحب العامة مثل MS Azure وGoogle BigQuery، ويتم تشغيلها فقط لمؤسســة واحــدة ذات عبء عمل المهام الحاســمة والمخاوف الأمنيــة، وهي توفر نفس الفوائد كخدمة عامة شــبيهة بالسحابة، والقابلية للتغيير، وتغيير موارد الحوسبة حسب الطلب، وما إلى ذلك. وتتمتع الشركات التي لديها سحابة خاصة بالتحكم المباشر على بياناتها وتطبيقاتها. وتكمن عيوب امتلاك ســحابة خاصة في تكلفة صيانة وإدارة الســحابة حيث يكون طاقم عمل تقنية المعلومات في الشركة هو المسؤول عن إدارتها.
- **سحابة عامة:** وفي هذا النموذج، يستخدم المشترك الموارد المعروضة من جانب مقدمي الخدمة على الإنترنت، وتتم إدارة البنية التحتية للسحابة بواسطة مقدم الخدمة. وتتمثل الميزة الأساسية لنموذج الســحابة العامة هذا في توفير وقت وتكلفة إعداد الأجهزة والبرامج المطلوبة لتنفيذ الأعمال. ومن أمثلة السحابة العامة منصة Microsoft Azure وGoogle App Engine وAmazon AWS.
- **ســحابة مختلطة:** تمنح الســحابة المختلطة الأعمال مرونة كبيرة بنقل أعباء العمل بين السحب الخاصة والعامة. فعلى ســبيل المثال، تستطيع أي شركة أن تستخدم التخزين السحابي المختلط

لتخزين بيانات المبيعات والتسـويق الخاصة بها، ومن ثَمّ اسـتخدام منصة سـحابة عامة مثل Amazon Redshift لتنفيذ الاستعلامات التحليلية لتحليل بياناتها. ويعد المطلب الرئيس هو الاتصال الشبكي وتوافق API (واجهة برمجة التطبيقات) بين السحابتين الخاصة والعامة.

كبار موردي منصة السحابة في التحليلات:

نحدد في هذا القسـم أولًا بعض المشـغلين الأساسـيين للسـحابة التي توفر البنية الأساسـية للتحليـلات كخدمة، بالإضافة إلى وظائف التحليلات المختـارة، ثم نذكر أيضًا عروض التحليلات - كخدمة التمثيلية التي قد تعمل حتى على منصات السحابة هذه.

Amazon Elastic Beanstalk: وهـي خدمـة مقدمة من خدمات ويب Amazon. وهي تسـتطيع نـشر وإدارة وتطوير تطبيقات الويـب وتدعم لغات البرمجة التاليـة: Java وRuby و Python وPHP وNET، عـلى خوادم مثل Apache HTTP وApache Tomcat وIIS. ويجب على المسـتخدم تنزيل رمز التطبيق، ويقـوم Elastic Beanstalk بنشر التطبيق، وموازنة العبء، والقياس التلقائي، ومراقبة صحة التطبيق. ولذا؛ يستطيع المستخدم التركيز على بناء مواقع الويب، وتطبيقات الجوال، وخلفية API، وأنظمة إدارة المحتوى، وSaaS، وما إلى ذلك، في حين يركز Elastic Beanstalk على التطبيقات والبنية التحتية لإدارتها. كما يمكن للمستخدم استخدام خدمات ويب Amazon أو بيئة تطوير متكاملة مثل Eclipse أو Visual Studio لتنزيل تطبيقهم. ويجب أيضًا على المستخدم دفع تكاليف موارد AWS اللازمة لتخزين وتنفيذ التطبيقات.

IBM Bluemix: وهو منصة سحابة تتيح للمستخدم إنشاء تطبيقات باستخدام عدة تقنيات مجانية على الحاسـب، كما يُمكِّن المسـتخدمين من نشر وإدارة التطبيقات المختلطة باسـتخدام البرنامج. ومع IBM Watson، والتي تتوفر خدماتها على IBM Bluemix، يستطيع المستخدمون الآن إنشـاء تطبيقات إدراكية للجيل القادم والتي تستطيع الاكتشاف، والابتكار، واتخاذ القرارات. كما يمكن اسـتخدام خدمات IBM Watson لتحليل العواطف وتحويل النص إلى كلام مسـموع بصوت طبيعي، حيث يسـتخدم Watson مفهوم الحوسـبة الإدراكية لتحليل النصوص ومقاطع الفيديو والصور، وهو يدعم لغات البرمجة مثل Java وGo وPHP وRuby وPython.

Microsoft Azure: قامت مايكروسوفت (Microsoft) بإنشاء منصة سحابية تسمى Azure لتقوم بإنشـاء ونشر وإدارة التطبيقات والخدمات من خلال شـبكة مراكز بيانات مايكروسوفت (Microsoft)، وهي تعمل على شكل كل من PaaS وIaaS وتقدم العديد من الحلول كالتحليلات ومستودعات البيانات والمراقبة عن بُعد والصيانة التنبؤية.

Google App Engine: وهو منصة حوسبة سحابية من خدمات Google تُستخدم لتطوير واستضافة التطبيقات، وهي تُدار من خلال مراكز بيانات Google، تدعم تطبيقات التطوير في لغات برمجة Python وJava وRuby وPHP، وتوفر بيئة BigQuery خدمات مستودع البيانات من خلال السحابة.

OpenShift: وهوعبـارة عن منصة تطبيقات سـحابية تابعة لـ Red Hat قائمة على نموذج PaaS. ومـن خلال هذا النموذج، يتمكن مطورو التطبيقات من نشر تطبيقاتهم على السـحابة. ويتوفـر نموذجـان مختلفان لــ OpenShift. يعمل أحدهما بمثابـة PaaS العامة والآخر بمثابة PaaS الخاص. إن OpenShift Online هونظام PaaS العام في Red Hat والذي يعرض تطوير وبناء واستضافة ونشر التطبيقات في السحابة. ويسمح PaaS الخاص، OpenShift Enterprise، بتطوير وبناء ونشر التطبيقات على خادم داخلي أو منصة سحابية خاصة.

التحليلات كخدمة (AaaS):

تعتـبر التحليـلات والحلول الإدارية القائمة على البيانات بمثابة التطبيقات التي تسـتعلم عن البيانـات لاسـتخدامها في تخطيط الأعمال، وحل المشـكلات، ودعم القرار، وهـي تتطور بسرعة وتُستخدم تقريبًا من جانب جميع المنظمات. ويتم إغراق المؤسسات بالمعلومات، ويُمثل الحصول عـلى أفـكار من هذه البيانـات تحديًا كبيرًا لهـا. وبالإضافة لذلك، توجد تحديـات تتعلق بأمن وجودة البيانات والالتزام. ويعتبر AaaS منصة تحليلية قابلة للتوسـع باستخدام نموذج التوصيل القائم على السـحابة، حيث تقوم أدوات تحليل بيانات ذكاء الأعمال والبيانات المختلفة بمساعدة الشركات في صنع القرار بشـكل أفضل والحصول على أفكار من بياناتها الضخمة. وتغطي المنصة جميـع النواحي الوظيفية ابتداءً من جمع البيانـات من الأجهزة المادية وانتهاءً بتصوير البيانات. ويوفـر AaaS للـشركات نموذجًا ذكيًا لإعداد التقارير والتحليلات مـما يمكنهم من التركيز على ما يفعلـون بأفضل ما يمكن. كما يمكن للعملاء تنفيذ تطبيقاتهم التحليلية الخاصة في السـحابة أو وضع بياناتهم على السحابة واستقبال أفكار مفيدة.

وتضم AaaS جوانب الحوسبة السحابية إلى تحليلات البيانات الضخمة، وتُمكِّن علماء البيانات والمحللين من خلال السماح لهم بالوصول إلى مجموعات بيانات المعلومات التي تتم إدارتها بشكل مركزي. ويمكنهم الآن استكشـاف مجموعات بيانات المعلومات بصورة تفاعلية أكثر واكتشـاف الأفـكار الأكثر ثراءً بشـكل أسرع، وبالتالي تجنب العديد من التأخـيرات التي قد يواجهونها أثناء اكتشـاف اتجاهات البيانات. فمثلًا، قد يوفر المورد الوصول إلى منصة تحليلية عن بُعد في مقابل دفع رسوم، ويتيح ذلك للعميل استخدام برامج التحليلات كلما لزم الأمر. ويعتبر AaaS جزءًا من

SaaS وPaaS وIaaS، وبالتالي يساعد تقنية المعلومات بشكل معنوي على تقليل كل من التكاليف وخطر الالتزام، في حين يرفع من إنتاجية المستخدمين.

على سـبيل المثال، يصل موظفــو eBay إلى شريحة افتراضية لخادم مســتودع البيانات الرئيس بحيث يمكنهم تخزين وتحليل مجموعات البيانات الخاصة بهم. وقد حققت أسواق بيانات eBay الافتراضية الخاصة نجاحًا كبيرًا فيما يتراوح من ٥٠ إلى ١٠٠ عملية في وقت واحد. وقد ألغت أسواق البيانــات الافتراضية احتياج الشركة إلى أســواق بيانات مادية جديدة، والتــي تقدر تكلفتها بمليون دولار لكل منها وتتطلب اهتمامًا طوال الوقت من عدة موظفين ذوي مهارات (Winter, 2008).

لــدى AaaS في الســحابة وفــورات الحجم والمدى من خــلال توفير العديد مــن التطبيقات التحليليــة الافتراضية مع مزيد من القابلية للتطوير وتوفــير التكلفة. ومع تزايد أحجام البيانات وعشرات من التطبيقات التحليلية الافتراضية، تكون هناك فرص بأن الكثير منها يستفيد بالمعالجة في أوقات مختلفة، وأنماط الاستخدام، والتكرارات (Kalakota, 2011).

يعتبر التنقيب في كل من البيانات والنصوص تطبيقًا آخر واعدًا جدًا من AaaS. يمكن أيضًا استخدام الإمكانيات التي يجلبها توجيه الخدمة (بجانب الحوسبة السحابية والموارد المجمّعة والمعالجة المتوازية) إلى عالم التحليلات من أجل التحســين واسع النطاق، ومشــاكل القرارات متعددة المعايير المعقدة جدًا، ونماذج المحاكاة الموزعة، ثم نحدد بعد ذلك عروض التحليلات المنتقاة والقائمة على السحابة.

التحليلات الممثِلة كعروض خدمة:

تحليلات ASTER كخدمة: تعتبر Teradata Aster هي المحرك الرئيس في تقديم التحليلات كخدمة. وتشــتمل Aster على كل من Aster MapReduce Analytics Foundation (تسمح بمعالجة البيانات عبر مجموعــات هائلة من البيانات)، وAster Graph Analytics مع كثير من أدوات ذكاء الأعمال الأخرى. ومع Aster AaaS، تستطيع الشركات الحصول على أفكار قيّمة من بياناتها، مما يساعدها في اتخاذ قرارات أفضل، دون عمل أي استثمار صريح في البنية التحتية.

IBM WATSON ANALYTICS: توفــر IBM جميع عروض التحليلات من خلال خدمة الســحابة Bluemix الخاصة بها؛ حيث يقــوم IBM Watson Analytics بدمج معظم خصائص وإمكانيــات التحليــلات التي يمكن بناؤها ونشرها من خــلال Bluemix. بالإضافة إلى ذلك، قدم IBM Watson Cognitive عرضًا أساســيًا قائمًا على السحابة يستخدم التنقيب في النص والتعلم العميق بمستوى عالٍ جدًا. وقد عرضنا لذلك من قبل في سياق الحديث عن التنقيب في النص.

MINEMYTEXT.COM: يعـد التنقيـب في النص واحـدًا من المجـالات النامية في عالم التحليلات. يعمل التنقيب في النص على تحديد موضوعات الوثائق ذات المستوى العالي، واستنتاج المشـاعر من الاسـتعراضات، وتصوير الوثيقة أو علاقات المصطلح / المفهـوم، وقد تم شرح ذلك بالتفصيل في فصل التنقيب في النص. وتقدم MineMyText.com هذه الإمكانيات في السـحابة من خلال موقع الويب الخاص بهم.

تحليـلات وإحصاءات SAS المرئيـة: يوفر معهد SAS برمجيـات التحليلات الخاصة به عند الطلب من خلال السـحابة. وحاليًا، تتوفر إحصاءات SAS المرئية فقط كخدمة سـحابة وتكون منافِسة لـ Tableau.

TABLEAU: وهو أحد أهم البرامج المرئية وقد تم تقديمه في سياق الحديث عن التحليلات الوصفية، ويتوفر أيضًا من خلال السحابة.

SNOWFLAKE: وهو أحد حلول مستودعات البيانات القائمة على السحابة؛ حيث يستطيع المستخدمون جمع بياناتهم من مصادر متعددة كمصدر واحد وتحليلها باستخدام Snowflake.

PREDIX BY GENERAL ELECTRIC: كـما ذكـرت التقارير، أن جنرال إلكتريك تركز على تطوير منصة إنترنت الأشـياء للتحليلات التنبؤية لمسـاعدة عملائها في الوصول إلى تشـغيل وإدارة المعـدات والمنصات الصناعية بشـكل أفضـل. وقد طورت الشركة عرضًـا تحليليًا جديدًا يطلق عليه Predix (وقد تمت مناقشته في حالة عملية ٨-٣)، وهو متوفر عبر السحابة من خلال خدمات Amazon Web Services وسيتوفر قريبًا من خلال Microsoft Azure.

تطبيقات تحليلات توضيحية باستخدام البنية التحتية للسحابة:

في هذا القسـم، نسلط الضوء على عدة تطبيقات تحليلات سحابية، وسنعرضهم كقسم واحد في مقابل حالات عملية منفردة.

مركز MD Anderson للسرطان يسـتخدم إمكانيات الحوسبة الإدراكية لشركة IBM Watson لمنح مرضى السرطان علاجًا أفضل:

يعتـبر مركز جامعـة Texas MD Anderson للسرطان (والـذي يطلق عليه مركز Anderson للسرطـان) واحدًا من أفضل مستشـفيات السرطـان في الولايات المتحدة. فمنذ عـام ١٩٤١، عالج مركز MD Anderson للسرطـان ٩٠٠ ألف مريض بالسرطان ولديه حوالي ١٩ ألف موظف.

حيــث يتم في كل عام علاج حــوالي ١٠٠ ألف مريض بالسرطان في MD Anderson ومحيطه المحلي والقومي. ونتيجة لذلك؛ تراكم لدى المركز الكثير من البيانات الإكلينيكية لــلأورام تخــص مرضى المركز، وتشــمل البيانــات الرعايــة المقدمة للمــرضى، والتجارب الإكلينيكية، ونتائج اختبارات المرضى، ويتم ذلك بشكل يومي، سواء في ملاحظات الباحثين والأطبــاء أو الموجودة في قواعد البيانات عن بُعد وملفات الأطباء الآخرين الذين عالجوا هؤلاء المرضى في الماضي. وقد أدرك مســؤولو MD Anderson أنه إذا تم جمع وتحليل كل هذه المجموعة من البيانات الضخمة في مصدر واحد وتحليلها، فســوف يساعد ذلك الأطباء في التجارب الإكلينيكية ومن ثَمّ تحديد أفضل اختيارات لعلاج مرضاهم.

وفي عام ٢٠١٢، أصدر مســؤولو MD Anderson برنامجًا بعنوان "Moon Shots" وقاموا بإجراء تحليلات لمحتوى IBM Watson للعثور على علاجات أفضل للسرطان. كما قاموا بتطوير منصة تقنية تســمى APOLLO تجمع كل البيانات المتفرقة غير المهيكلة للمــرضى في نظام واحد للســجلات الطبيــة الإلكترونية (EMR)، حيــث يتكامل IBM Watson بسلاســة مع نظام EMR ويســتخدم تحليلات المحتوى لإنشــاء ملف تعريف شامل لكل مريض بالسرطان في شكل مهيكل للأطباء. فهي تساعد الأطباء على التشخيص الأفضل لحالة المريض وتُمكّنهم من إجراء مقارنة بين المرضى استنادًا إلى مجموعة جديدة من السمات المعتمدة على البيانات. والآن، يستطيع فريق أطباء مركز MD Anderson مقارنة مجموعة من المرضى لتحديد أولئك الذين يســتجيبون بشكل مختلف للعلاجات واكتشــاف السمات المســؤولة عن هذه الاختلافات. كما يمكن للأطباء الآن عرض المرضى على المشاركة في التجارب الإكلينيكية على العلاجات الجديدة بناءً على الدليل والخبرة.

Sources: MDanderson.org. (2013). MD Anderson taps IBM Watson to power "Moon Shots" mission. https://www.mdanderson.org/newsroom/201310//md-anderson-ibm-watson-work-togetherto-fight-cancer.html (accessed August 2016); IBM.com (2015). Smarter care at MD Anderson. http://www-03.ibm.com/software/businesscasestudies/us/en/corp?synkey=H447240O66679Z38 (accessed August 2016); YouTube.com. (2014). Smarter care at MD Anderson. https://www.youtube.com/watch?v=savJ8VQ0kcA (accessed August 2016). Wikipedia.org. (n.d.). University of Texas MD Anderson Cancer Center. https://en.wikipedia.org/wiki/University_of_Texas_MD_Anderson_Cancer_Center (accessed August 2016).

مدارس التعليم العام في تاكوما، واشنطون تستخدم Microsoft Azure Machine Learning للتنبؤ بالتسربات من التعليم:

تعتبر مدارس تاكوما العامة هي الإدارة التعليمية الرئيسة في تاكوما، واشنطن. وتتكون من ٣٥ مدرسة ابتدائية و٩ مدارس متوسطة و٩ مدارس ثانوية. وهي تعتبر ثالث أكبر إدارة تعليمية في ولاية واشنطن، بأكثر من ٣٠٠٠٠ طالب وأكثر من ٥٠٠٠ موظف.

وقد أشارت دراسة أجريت في عام ٢٠٠٧ إلى خمس مدارس ثانوية من مدارس تاكوما العامة على أنها "مصانع للتسرب". وبحلول عام ٢٠١٠، كان الموقف في غاية السوء. فقد حصل فقط ٥٥٪ من طلاب المدارس الثانوية على شهادة الدبلوم الخاصة بهم في الوقت المحدد، وهومعدل أقل بكثير من المعدل الوطني البالغ ٨٨٪. كان لدى الإدارة التعليمية الكثير من البيانات الخاصة بالطلاب كدرجات الاختبار، وأماكن الإقامة، ومدارسهم الابتدائية والمتوسطة التي أتوا منها. فأرادوا باستخدام هذه البيانات أن يوجدوا حلًا لرفع معدلات التخرج. كما أرادوا أيضًا التنبؤ المسبق بالطلاب الذين من المحتمل تسربهم من التعليم، وذلك لمنحهم اهتمامًا خاصًا بمشاكل الطفل ومن ثَمّ اتخاذ إجراءات وقائية.

شرعت مدارس تاكوما العامة في استكشاف حلول متنوعة لذكاء الأعمال من أجل تحليل بياناتها. وتعاونت خدمات استشارات مايكروسوفت (Microsoft) مع المنطقة التعليمية وقامت بإنشاء مستودع بيانات يلتقط معلومات تخص الطالب كالصحة والدرجات ونسبة الحضور وتفاصيل إضافية مستقاة من أنظمة معلومات طلاب المدرسة. كما ساعدت خدمات الإكسيل التابعة لمايكروسوفت (Microsoft) وSharePoint المدرسين في مراجعة البيانات التاريخية لطلابهم والإجراءات التي تم اتخاذها على المقاييس المشاهدة. وقد ساعدهم ذلك على قياس تقدم الأطفال مجتمعين وتحديد مقدار جودة أداء المدرسة في مساعدة الأطفال على التقدم للأمام.

ثانيًا، أرادت الإدارة التعليمية التنبؤ بالطلاب المحتمل تسربهم من التعليم، حتى يتمكنوا من مساعدتهم بشكل مسبق والعمل معهم عن قرب. ومرةً أخرى، تعاون مسؤولو الإدارة مع شركة مايكروسوفت (Microsoft) لإنشاء نموذج بيانات يبرهن المفهوم ويستند على Microsoft's Azure Machine Learning (ML)، وهو حل تحليلي تنبؤي يستند إلى المنصة السحابية التابعة لمايكروسوفت (Microsoft). وقد قام هذا النموذج بتحليل البيانات التي تم رفعها إلى Azure من أنظمة معلومات متعددة على مباني الحرم الجامعي. وقام مصنع بيانات Azure بتمكين خط أنابيب

تنبؤي يســتخدم نموذج Azure ML للتنبؤ بتعرض الطالب لخطر التسرب. ويتم وضع النتائــج التنبؤيــة من خلال قاعــدة بيانــات Microsoft Azure SQL، بحيث يتمكن أعضــاء طاقم العمل وأعضــاء مجلس الإدارة التعليمية من مراجعة النتائج باســتخدام لوحــات معلومات Power BI. وهكذا تمكنت الإدارة من رفع معدلات التخرج لمدارس تاكوما العامة من ٥٥٪ في عام ٢٠١٠ إلى ٨٢,٦٪ في عام ٢٠١٦ وذلك بمساعدة التحليلات التنبؤية وMicrosoft Azure Machine Learning.

Sources: Blogs.technet.microsoft.com (2015). ML predicts school dropout risk & boosts graduation rates. https://blogs.technet.microsoft.com/machinelearning/201504/06// ml-predicts- school-dropout-risk-boosts-graduation-rates/(accessed August 2016); Customer.microsoft.com (2015). Tacoma Public Schools: Predicting student dropout risks, increasing graduation rates with cloud analytics. https://customers.microsoft.com/Pages/CustomerStory.aspx?recid=20703 (accessed August 2016); YouTube.com (2016). The saving power of data. https://www.youtube.com/watch? v=rfAoKs8XxzY (accessed August 2016).

مركز Dartmouth-Hitchcock الطبي يقدم الرعاية الصحية المســبقة الشخصية باستخدام Microsoft Cortana Analytics Suite:

يقــع مركز Dartmouth-Hitchcock الطبي (DHMC) في لبنان، نيوهامبشــاير. وهــو المركز الطبــي الأكاديمي الوحيد في نيوهامبشــاير وبه ما يقــرب من ٤٠٠ سرير. ويســعى هذا المركز أن تحدد بشــكل مســبق صحة الأفراد المحتمل إصابتهم بالأمراض وبالتــالي وقايتهم قبل الوقوع في فخ المرض. ويتمثــل هدفهم في توفير الرعاية الصحية الشــخصية بتكلفة أقل. فتقــوم الممرضات والمدربون الصحيــون بتتبع الحالة الصحية للمريض أولًا بأول، ويتم جمع البيانات باستخدام أجهزة استشعار مرفقة بأجهزة ضغط الدم وأجهزة قياس النبض وأجهزة تتبع النشــاطات مثل Microsoft Band. ويتم نقل هذه البيانات إلى ســحابة Azure باســتخدام الهواتف الذكية، ثــم يتم عرض البيانات على لوحات المعلومات التي تتم مراقبتها بشــكل مســتمر ٢٤ ساعة في اليوم سبعة أيام في الأسبوع بواسطة الممرضات المسجلات باستخدام Microsoft Cortana Analytics Suite. وعندما تتخطى بيانات المريض حد الســلامة، يتم تنبيه الممرضات، اللاتي يقمن فورًا بالتواصل مع المريض، وبالتالي يتم تجنب المريض لخطر أي مشكلة خطيرة. وكنتيجة لاستخدام تقنية مايكروسوفت (Microsoft) هذه، يمكن مراقبة صحة المريض عن بُعد من منزل المريض، مما يساعد على تقليل تكلفة الزيارات المنتظمة للأطباء.

Sources: Blogs.microsoft.com (2015). Dartmouth-Hitchcock ushers in a new age of proactive, personalized healthcare using Cortana Analytics Suite. http://blogs.microsoft.com/transform/201513/07//dartmouth-hitchcock-ushers-in-a-new- age-of-proactive-personalized-healthcare-using-cortana-ana- lytics-suite/(accessed August 2016); Enterprise.microsoft.com. (2015). How Dartmouth-Hitchcock is challenging healthcare's status quo with Cortana Analytics. https://enterprise.microsoft. com/en-us/industries/health/how-dartmouth-hitchcock-is-chal- lenging-healthcares-status-quo-with-cortana-analytics/(accessed August 2016); YouTube.com. (2015). Dartmouth-Hitchcock revolutionizes the U.S. healthcare system. https://www.youtube.com/watch?v=-wVeHZNn8aU (accessed August 2016).

Mankind Pharma تســتخدم بنيــة الســحابة التحتية IBM لتقليــص وقت تنفيذ التطبيق بنسبة ٩٨٪:

تعتــبر Mankind Pharma شركــة أدوية مقرها في نيودلهــي، الهند، هي رابع أكبر منتج للوصفات الطبية المخدِرة في الهند حيث يعمل بها ١١ ألف موظف وتبلغ إيراداتها ٦٠٠ مليون دولار. ومع أعمالها المتنامية، كانت Mankind Pharma تبحث عن بيئة اســتضافة ســحابية لتوفير بنية تحتية لمنصة الموارد البشرية (HR) الخاصة بها وللمهمات الحساسة الأخرى.

وفي ســبيل ذلك شرع الفريق الفني بشركة Mankind في اســتخدام منصة سحابة IBM وتســمى Softlayer، والتي تحتوي على مراكــز بيانات وخوادم فلزية معزولة في جميع أنحاء العالم، مما يساعد Mankind على القيام بمهام أعمالها الحساسة على نطاقٍ عالمي. وبالفعل انخفض وقت تنفيذ تطبيق الشركة بنسبة ٩٨٪ وذلك باستخدام خدمة البنيــة التحتية لــــ Softlayer. والآن أصبح بإمكان العملاء الحصــول على التطبيق في غضون ساعات بدلًا من ١٥ يومًا كما كان يحدث في السابق.

Sources: IBM.com (2014). Softlayer hosting platform reduces application implementation time by 98%. http://www-03.ibm.com/software/businesscasestudies/us/en/corp?synkey=Y979749I50926G25 (accessed August 2016); CIO.in. (2015). Mankind Pharma finds an antidote in IBM solution for improving app implementation time (accessed August 2016). http://www.cio.in/solution-center/emc/55281; CxOtoday.com (2014). Cloud platform to help Pharma Co accelerate growth. http://www.cxotoday.com/story/mankind-pharma-to-drive-growth-with-softlayers-cloud-platform/(accessed August 2016); Wikipedia.org.(n.d.). Mankind Pharma. https://en.wikipedia.org/wiki/Mankind_Pharma (accessed August 2016);

Mankindpharma.com. (n.d.). Overview. https://www.mankindpharma.com/company/companyoverview (accessed August 2016).

الخليج للطيران تستخدم البيانات الضخمة للحصول على رؤية أعمق للعملاء:

تعتبر الخليج للطيران هي شركة النقل الوطني لدولة البحرين. وهي شركة طيران دولية رائدة يعمل بها ٣٠٠٠ موظف، وتخدم ٤٥ مدينة في ٢٤ دولة عبر ٣ قارات. وهي رائدة في توفير الضيافة العربية التقليدية للعملاء. ولمعرفة المزيد عما يشعر به عملاؤهم بشأن خدمات ضيافتهم، تابعت الشركة ما يشاركه عملاؤهم على وسائل التواصل الاجتماعي في هذا الصدد. وكان التحدي هو تحليل جميع تعليقات ومنشورات العملاء، فقد كان هناك مئات الآلاف من المشاركات كل يوم. وتعتبر مراقبة هذه المشاركات يدويًا مهمة شاقة ومستهلكة للوقت كما ستكون عرضة للخطأ البشري.

أرادت شركة الخليج للطيران جعل هذه المهمة وتحليل البيانات لمعرفة الاتجاهات السوقية حديثة الظهور تتم بصورة تلقائية. كما أرادت الشركة تأسيس بنية تحتية قوية لاستضافة مثل هذا الحل لمراقبة وسائل التواصل الاجتماعي والتي ستكون متاحة على مدار الساعة وخفيفة الحركة عبر الحدود الجغرافية.

طوّرت الخليج للطيران حلًا لتحليل المشاعر، "تحليل المشاعر العربية"، والذي يحلل منشورات وسائل التواصل الاجتماعي الإنجليزية والعربية. وتستند أداة تحليل المشاعر العربية على توزيع Cloudera لإطار عمل Hadoop للبيانات الضخمة. وهي تعمل على بيئة السحابة الخاصة التابعة لشركة الخليج للطيران كما تستخدم منصة Red Hat JBoss Enterprise Application. تحمل هذه السحابة الخاصة حوالي ٥٠ تيرابايت من البيانات، ويمكن لأداة تحليل المشاعر العربية تحليل آلاف المنشورات على وسائل التواصل الاجتماعي، مما يوفر نتائج المشاعر في ظرف دقائق.

وقد حققت شركة الخليج للطيران وفورات جوهرية في التكاليف من خلال وضع تطبيق "تحليل المشاعر العربية" على بيئة السحابة الخاصة الحالية للشركة حيث إنهم لم يكونوا بحاجة لأن يستثمروا في إعداد البنية التحتية لنشر التطبيق. وتستعين الخليج للطيران بأداة "تحليل المشاعر العربية" في تحديد العروض الترويجية لعملائها على أساس زمني وتساعدهم على الجلوس في مقدمة منافسيها. ولمواجهة حالة تعطل الخادم الأساسي، أنشأت الشركة "صورًا شبحية" للخادم يمكن نشرها بسرعة، ويمكن أن تبدأ الصورة في لعب دور الخادم. ويقوم حل البيانات الضخمة بسرعة وكفاءة بالتقاط

المنشورات بشكل دوري وتحويلها إلى تقارير، مما يجعل شركة الخليج للطيران على دراية تامة بأي تغييرات في المشاعر أو تغيرات في الطلب، مما يمكنها من تقديم استجابة سريعة. فكان لأفكار حل البيانات الضخمة أثرًا إيجابيًّا على أداء موظفي شركة الخليج للطيران.

Sources: RedHat.com. (2016). Gulf Air builds private cloud for Big Data innovation with Red Hat Technologies. https://www.redhat.com/en/about/press-releases/gulf-air-builds-private-cloud-Big-data-innovation-red-hat-technologies; (accessed August 2016); RedHat.com (2016). Gulf Air's Big Data innovation delivers deeper customer insight. https://www.redhat.com/en/success-stories(accessed August 2016); ComputerWeekly.com. (2016). Big-data and open source cloud technology help Gulf Air pin down customer sen- timent. http://www.computerweekly.com/news/450297404/Big- data-and-open-source-cloud-technology-help-Gulf-Air-pin-down- customer-sentiment (accessed August 2016).

Chime تحسن تجربة العملاء باستخدام Snowflake:

يوفـر Chime، وهوخيار مـصرفي، بطاقة ائتمان فيـزا، وحسـاب FDIC المؤمن للصرف والتوفير، وتطبيق على الهاتف الجوال يسهل الخدمات المصرفية للأفراد. وأراد مسؤولو Chime التعرف على مشـاركات عملائه. كما أرادوا تحليـل البيانات عبر المنصات الخاصة بهم للجوّال، والويـب، والنهاية الخلفية لتحسـين تجربة المسـتخدم. ومع ذلك، فقد كانت عملية سـحب البيانات وتجميعها من مصادر متعددة مثل خدمات إعلانات Facebook وGoogle والأحداث من الأدوات التحليلية الأخرى التابعة لطرف ثالث مثل JSON (JASON Object Notation)، مهمـة مرهقة. لقد سـعوا للوصـول إلى حل يُمكِّنهم مـن تجميع البيانات مـن هذه المصادر المتعـددة وتحليل مجموعة البيانات. لقـد كان Chime بحاجة إلى حل يمكنه معالجة مصادر بيانات JSON والاستعلام عنها باستخدام جداول قاعدة بيانات SQL المعيارية.

شرع Chime في اسـتخدام حل مسـتودع بيانـات Snowflake Elastic؛ إذ قامت Snowflake بسـحب البيانـات مـن جميع مصـادر بيانـات Chime والبالغ عددها ١٤ مصـدرًا، وقد شـمل ذلك بيانـات مثل وثائق JSON مـن التطبيقات. فقد عملت Snowflake على مسـاعدة Chime في تحليل بيانات JSON بسرعة لتحسـين خدمات الأعضاء وتقديم تجربة مصرفية أكثر خصوصية للعملاء.

Source: Adapted from Snowflake.net. (n.d.). Chime delivers personalized customer experience using Chime. http://www.snow-flake.net/product (accessed August 2016).

إننــا بصدد دخــول "عصر البيتابايــت"، وتبدأ الأســاليب التقليدية للبيانــات والتحليلات في عرض حدودها. وتعتبر تحليلات الســحابة حلًا بديلًا حديث الظهور لتحليل البيانات على نطاق واســع. وتشــتمل الأنظمة الســحابية الموجهة بالبيانات على التخزين والحوسبة في بيئة موزعة وافتراضية. وتعتبر الميزة الأساســية لهذه العروض هي الانتشار السريع لأدوات التحليل المتقدمة بين المســتخدمين، دون استثمار معنوي في الاســتحواذ على التقنية. غير أن هذه الحلول يصاحبها أيضًــا العديد مــن التحديات، مثل الأمان ومســتوى الخدمة والتحكم في البيانــات. فقد اكتنف الحوسبة السحابية عدد من المخاوف، منها فقدان السيطرة والخصوصية، والمسؤوليات القانونية، والقضايا السياسية عبر الحدود، وما إلى ذلك. ووفقًا لتحالف Cloud Security، فإن أخطر ثلاثة تهديدات أمنية في الســحابة هي فقدان البيانات وتسربهــا، وتعطل الأجهزة والمعدات، والواجهة غير الآمنة. فجميع البيانات في الســحابة يســتطيع مقدم الخدمة الوصــول إليها، وبالتالي يمكنه أن يغير البيانات بشــكل غير مقصود أو متعمد كما يمكنه تمرير البيانات إلى طرف ثالث لأغراض قانونية دون طلب من الشركة. ولا تزال البحوث محدودة في هذا المجال. وكنتيجة لذلك، فهناك فرصة كبيرة لإدخال النمذجة التحليلية والحســابية والمفاهيمية إلى ســياق علم الخدمة، وتوجيه الخدمة، والذكاء الســحابي. ومع ذلك، تعتبر الحوســبة السحابية مبادرة مهمة لمحترف التحليلات أن يشاهدها على أنها مجال سريع النمو.

أسئلة مراجعة على القسم ٨-٣:

١- عرِّف الحوسبة السحابية. وكيف تتعلق بـ PaaS وSaaS وIaaS؟

٢- أعطِ أمثلة لشركات تقدم الخدمات السحابية.

٣- كيف تؤثر الحوسبة السحابية على ذكاء الأعمال؟

٤- كيف يقوم DaaS بتغيير طريقة التعامل مع البيانات؟

٥- ما هي الأنواع المختلفة للمنصات السحابية؟

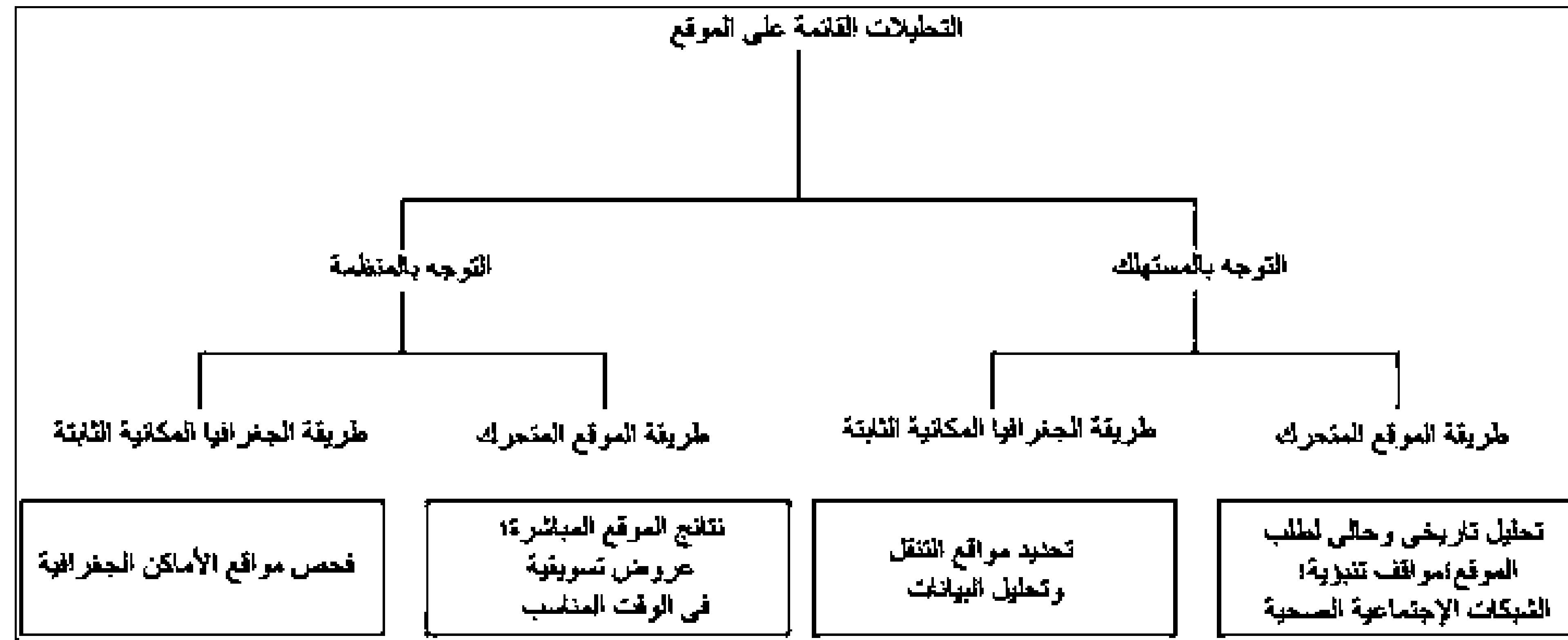
٦- لماذا تعتبر AaaS فعالة من حيث التكلفة؟

٧- اذكر على الأقل أسماء ثلاثة من كبار مقدمي الخدمات السحابية.

٨- أعطِ على الأقل ثلاثة أمثلة لمقدمي خدمات AaaS.

٨-٤ التحليلات القائمة على الموقع بالنسبة للمنظمات:

لقــد رأينا حتى الآن، العديد مــن الأمثلة على المنظمات التي تســتخدم تقنيات تحليلية من أجل اكتســاب نظرة ثاقبة في عملياتها الحالية من خــلال تقديم المعلومات، والتحليلات التنبؤية، والتوقع، وأســاليب الأمثلية. وفي هذا القسم، ســوف نتعرف على اتجاه ناشئ بالغ الأهمية وهو دمج بيانات الموقع في التحليلات. ويُعطي الشكل (٨-٦) تصنيفًا للتطبيقات التحليلية القائمة على الموقع. وســوف نقوم أولًا باســتعراض التطبيقات التي تســتفيد من بيانات الموقع الثابتة والتي عادةً ما تُســمى بالبيانات الجغرافية المكانية. ومن ثَمَّ ســنقوم بعد ذلك بفحص التطبيقات التي تســتفيد من جميع بيانات الموقع الذي يتم إنشاؤه بواسطة أجهزة اليوم. ويركز هذا القسم أولًا على تطبيقات التحليلات التي تطورها المنظمات من أجل اتخاذ أفضل قرارات في إدارة العمليات، واســتهداف العملاء، والتقدم، وهكذا. ثم سنقوم أيضًا باستكشاف تطبيقات التحليلات التي يتم تطويرها ليستخدمها المستهلك بشكل مباشر، كما أن بعضها يستفيد أيضًا من بيانات الموقع.



شكل ٨-٦: تصنيف تطبيقات التحليلات القائمة على الموقع

التحليلات الجغرافية المكانية:

عادةً ما يتم تمثيل رؤية موحدة للأداء العام للمنظمة من خلال أدوات التصوير المرئي التي توفر معلومــات قابلة للتطبيق. وقد تتضمن المعلومات القيم الحالية والقيم المتوقعة للعوامل التجارية المختلفــة ومؤشرات الأداء الرئيســة (KPIs). إن النظر في مؤشرات الأداء الرئيســة كأرقام إجمالية عبر كل من الرســوم البيانية والمخططات المختلفة يمكن أن يكون ساحقًا. وهناك خطر كبير يتعلق بفقدان فرص النمو المحتملة أو عدم تحديد مجالات الجدل. وتستخدم المنظمات الخرائط البصرية التــي يتم تعيينها جغرافيًّا والتــي تعتمد على بيانات الموقع التقليدية، كبديل لمجرد عرض التقارير، وعادةً ما يتم تجميع هذه الخرائط البصرية بواســطة الرموز البريدية. وتعمل الأســاليب التحليلية

القائمة على الموقع التقليدي والتي تســتخدم الترميز الجغرافي للمواقع التنظيمية والمســتهلكين على عرقلة المنظمات عن فهم آثار «الموقع الحقيقي». وتقدم المواقع القائمة على الرموز البريدية عرضًا مجمعًــا لمنطقة جغرافية كبيرة. وهذه الدقة المنخفضة قد لا تســاعد في تحديد فرص النمو داخل المنطقة، حيث يمكن أن يتغير موقع العملاء المســتهدفين بسرعة. وهكذا، فإن الحملات الترويجية الخاصة بالمنظمة قد لا تســتهدف العملاء المناســبين إذا اعتمدت على الرمــوز البريدية. ولمعالجة هذه المخاوف، فإن المنظمات تضم الموقع والتوســعات المكانية للتحليلات (Gnau, 2010). حيث إن إضافة مكونات الموقع التي تعتمد على السمات الطولية والعرضية للأساليب التحليلية التقليدية تُمكِّن المنظمات من إضافة بعد جديد يتعلق بالسؤال «أين» إلى تحليلات الأعمال التقليدية الخاصة بهم، والذي يُجيب حاليًا على الأسئلة التي تتعلق بـ«من» و«ماذا» و«متى» و«ما مقدار».

الجدير بالذكر أن البيانات التي تعتمد على الموقع يمكن الحصول عليها الآن بسهولة من نظم المعلومات الجغرافية (GIS)، وهي تُســتَخْدَم لرصــد، وتخزين، وتحليل، وإدارة البيانات المرتبطة بموقع باستخدام تقنيات الاستشعار المتكاملة، والأنظمة العالمية لتحديد المواقع المثبتة في الهواتف الذكية، أو من خلال نشر تقنية تحديد الهوية باســتخدام موجات الراديو(RFID) في مجال البيع بالتجزئة وصناعات الرعاية الصحية.

ومــن خــلال دمج المعلومات التي تدور حــول الموقع مع بيانات الأعــمال المهمة الأخرى، تقوم المنظــمات عندئذٍ بإنشــاء معلومات للمواقــع (Krivda, 2010). وتقوم ميــزة ذكاء الموقع بتمكين المنظــمات من اكتســاب رؤى دقيقة واتخــاذ قرارات أفضل من خلال تحســين كلٍ مــن العمليات والتطبيقات المهمة. وتعمل المنظمات الآن على إنشاء خرائط تفاعلية والتي من شأنها دعم التعمق في التفاصيــل حول أي موقع، فضلًا عن أنها توفر للمحللين القدرة على البحث عن اتجاهات جديدة والربط بين العوامل الخاصة بالموقع عبر مؤشرات الأداء الرئيسة المتعددة KPI. ويمكن للمحللين الآن تحديد الاتجاهات والأنماط بدقة في كلٍ من الإيرادات، والمبيعات، والربحية عبر المناطق الجغرافية.

ومــن خلال دمــج التفاصيــل الديموغرافية داخــل المواقع، يســتطيع تجــار التجزئة تحديد كيفية اختلاف المبيعات حســب مســتوى الســكان والقرب من المنافســين الآخرين؛ بالإضافة إلى أنهم يســتطيعون تقييم الطلب وكفاءة عمليات سلســلة التوريد. كما تستطيع شركات المنتجات الاســتهلاكية تحديد الاحتياجات الخاصة للعملاء ومواقع شكاوى العملاء وتتبعهم بسهولة وصولًا إلى المنتجات. ويمكن لمندوبي المبيعات تحقيق أهدافهم البيعية من خلال تحليل جغرافيتهم.

وتُعد شركة ESRI (esri.com) هي الشركة الرائدة في السوق في توفير بيانات أنظمة المعلومات الجغرافيــة؛ حيث تقوم ESRI بترخيــص برنامج ArcGIS الخاص بهــا لآلاف العملاء بما في ذلك

التجاريون، والحكومة، والجيش. وقد يستغرق الأمر كتابًا أو أكثر لتوضيح تطبيقات قاعدة بيانات أنظمة المعلومــات الجغرافية لشركة ESRI وبرامجها! وهناك شركة أخرى وهي grindgis.com تحــدد أكثر من ٦٠ فئة من تطبيقــات أنظمة المعلومات الجغرافيــة/http://grindgis.com). GIS (blog/gis-applications-usesوهناك بعض الأمثلة التي لم نذكرها بعد وتشمل ما يلي:

- التطبيقات الزراعية: فمن خلال الجمع بين كل من الموقع، والطقس، والتربة، والبيانات المرتبطة بالمحاصيل، يمكن التخطيط بدقة عالية لتطبيقات الري والأســمدة. وتتضمن الأمثلة على ذلك بعــض الشركات مثل شركة sstsoftware.com وشركــة sensefly.com (فكل منهما تجمع بين أنظمــة المعلومات الجغرافية GIS وأحدث المعلومات التــي جُمِعَت من خلال طائرات بدون طيار وتقنيات أخرى ناشئة).
- تحليــل الجريمــة: إن ربط بيانات الجريمة بما في ذلك تاريخ الجريمة، ووقتها، ونوعها مع بيانات أنظمة المعلومات الجغرافية يوفر رؤى ثاقبة لأنماط الجريمة وموظفي الشرطة.
- التنبــؤ بانتشــار المرض: من أوائل أمثلــة التحليلات الوصفية المعروفة تحليــل تفشي وباء الكوليرا في لنــدن عــام ١٨٥٤. خطط الدكتور John Snow حالات الكوليرا على خريطة واســتطاع دحض النظرية التي ترى أن تفشي وباء الكوليرا ســببه ســوء الهواء. فقد ساعدته الخريطة على أن يحدد بدقة أن سبب التفشي يرجع إلى بئر ماء سيئة. (TheGuardian.com, 2013). وقد قطعنا شوطًا كبــيرًا مــن الحاجة إلى تخطيط الخرائط يدويًــا، غير أن الفكرة المتعلقة بالقــدرة على تتبع تفشي الأمراض ومن ثَم التنبؤ بها، مثل الأنفلونزا، باستخدام أنظمة المعلومات الجغرافية GIS والبيانات الأخــرى أصبحــت مجالًا رائدًا في حد ذاته. وقد قدم الفصل الســابع مثالًا على اســتخدام بيانات وسائل التواصل الاجتماعي مع بيانات أنظمة المعلومات الجغرافية لتحديد اتجاهات الأنفلونزا.

وبالإضافة إلى ذلك، تستطيع المنظمات باستخدام ذكاء الموقع، أن تقوم بتغطية الطقس والآثار البيئية بشكل سريع وتوقع مستوى التأثير على عمليات الأعمال المهمة. ومع التقدم التقني، يتم الآن إدراج البيانات الجغرافية المكانية مباشرةً في مســتودعات بيانات المؤسســة، وتقوم تحليلات قاعدة البيانات التي تعتمد على الموقع بتمكين المنظمات من إجراء العمليات الحســابية المُعقدة بكفــاءة عالية والحصول عــلى رؤية واحدة لجميع البيانات الموجهة مكانيًّا، وكشــف الاتجاهات المختبئــة والفــرص الجديدة. فمثــلًا، يدعم مســتودع البيانات الخاص بـ تيراداتــا ميزة البيانات الجغرافيــة المكانية معتمدًا عــلى المعيار SQL/MM. ويتم رصد ميــزة الجغرافيا المكانية كنوع بيانات هندسي جديد يُســمى ST_GEOMETRY، وهو يدعم مجموعة كبيرة من الأشكال من

نقاط بسيطة، وخطوط، وينحني إلى المضلعات المعقدة في تمثيل المناطق الجغرافية، وهم يقومون بتحويـل البيانـات غير المتداخلة الخاصة بمواقع أعمالهم التشـغيلية من خـلال دمج إحداثيات خطـوط الطول والعرض. ويتم دعم هـذه العملية الخاصة بالترميز الجغرافي بسـهولة من قِبل شركات الخدمـات مثل شركة NAVTEQ، وشركـة Tele Atlas، والتي تحافظ على قواعد بيانات عالمية للعناوين ذات الخصائص الجغرافية المكانية والاسـتفادة من أدوات تنظيف العناوين مثل Informatica وTrillium، واللتان تدعمان رسـم خرائط الإحداثيات المكانية للعناوين كجزء من وظائف الاستخراج، والتحويل، والتنزيل.

وتقوم المنظمات عـبر مجموعة متنوعة من قطاعات الأعمال باسـتخدام التحليلات الجغرافية المكانية. وسنقوم فيما يلي باستعراض بعض الأمثلة؛ حيث تقدم حالة عملية (٨-٤) مثالًا على كيفية اسـتخدام المعلومات التي تسـتند إلى الموقع في اتخاذ قرارات اختيار الموقع في توسيع رقعة الشركة. وتوضح حالة عملية (٨-٥) تطبيقًا آخر والذي يتجاوز مجرد قرار الموقع.

حالة عملية ٨-٤

Great Clips تستخدم التحليلات المكانية لتوفير الوقت في قرارات الموقع

يعد Great Clips أكبر وأسرع صالونات الشعر في العالم نموًا، وله أكثر من ٣٠٠٠ فرع ما بين الولايات المتحدة الأمريكية وكندا. يعتمد نجاح Great Clips في منح حق الامتياز للغير على إستراتيجية نمو قائمة على افتتاح فروع جديدة بشكل سريع في المواقع والأسواق المناسبة. كانت Great Clips بحاجة إلى تحليل المواقع بناءً على متطلبات قاعدة العملاء المحتملـة، والاتجاهـات الديموغرافية، وأثر المبيعات على حقـوق الامتياز القائمة بالفعل في الموقـع المسـتهدف. ومن الأهمية بمكان في هـذا الصدد اختيار الموقـع الجيد. وقد اسـتغرقت العمليـات الحالية وقتًا طويـلًا لتحليل موقع واحد وعـدد هائل من العمالة يتطلب موارد تحليل مكثفة لتقييم البيانات بشكل يدوي من مصادر البيانات المتعددة.

ومـع آلاف المواقع التي يتـم تحليلها في كل عام، كان التأخير ينذر بخطر خسـارة المواقع الأولية لحسـاب المنافسـين كـما كان يبرهن على تكلفتها الباهظة؛ فاسـتعانت Great Clips بمقاولين خارجيين للتغلب على مشـكلة التأخير. وقامت الشركة بإنشـاء تطبيق سـير عمل اختيـار الموقع لتقييم أماكن الفروع الجديدة باسـتخدام الإمكانيات التحليليـة للجغرافيا المكانية التابعة لشركة Alteryx. وأي موقع جديد يتم تقييمه من خـلال وقت القيادة التقريبية إليه وسـهولة خدمة جميع العملاء الموجودين في محيط

شبكة Great Clips. كما مكّن الحل القائم على Alteryx في تقييم كل موقع جديد بناءً على البيانات الديموغرافية والسلوكية للمستهلكين، والتوافق مع ملفات تعريف عملاء Great Clips الحاليين والأثر المحتمل لإيرادات الموقع الجديد على المواقع الموجودة من قبل. وكنتيجة لاستخدام فنيات التحليل القائمة على الموقع، استطاعت Great Clips تقليل وقت تقييم المواقع الجديدة بنسبة ٩٥٪ تقريبًا، وتم إجراء التحليل المكثف للعمالة بصورة آلية وتم تطويره في تحليل جمع البيانات، ورسم الخرائط، وتطبيق إعداد التقارير التي يمكن استخدامها بسهولة من قِبَل مديري العقارات غير الفنيين. بالإضافة إلى ذلك، تمكنت الشركة من تنفيذ تحليلات تنبؤية مسبقة لأي موقع جديد، بحيث لا تستغرق العملية بأكملها الآن سوى بضع دقائق.

أسئلة للمناقشة:

١- كيف يتم استخدام تحليلات الجغرافيا المكانية في Great Clips؟

٢- ما هي المعايير التي ينبغي على الشركة مراعاتها عند تقييم أماكن المواقع الجديدة في المستقبل؟

٣- هل تستطيع التفكير في تطبيقات أخرى قد تستفيد من بيانات الجغرافيا المكانية؟

Source: Adapted from Alteryx.com. (n.d.). Great Clips. alteryx.com/sites/default/files/resources/files/case-study-great-chips.pdf (accessed August 2016).

حالة عملية ٨-٥

ستاربكس تستغل نظم المعلومات الجغرافية (GIS) وتحليلاتها لتنمو في كل أنحاء العالم

يعتبر أحد أهم التحديات التي تواجه أي منظمة تسعى للتوسع هو تحديد موقع فرعها التالي، وهذا ما تواجهه ستاربكس. ولتحديد مواقع فروع جديدة، يقوم أكثر من ٧٠٠ موظف في ستاربكس (يشار إليهم كشركاء) في ١٥ دولة باستخدام تخطيط السوق القائم على ArcGIS وحل ذكاء الأعمال الذي يسمى Atlas؛ حيث تمد Atlas الشركاء بمهام سير العمل والتحليل ومعلومات عن أداء الفرع بحيث يمكن للشركاء المحليين في المجال اتخاذ القرارات عند تحديد فرص أعمال جديدة.

وكما ذكرت تقارير مصادر متعددة، يستخدم صانعو القرار المحليين Atlas لفهم الاتجاهات السكانية والطلب. فيوجد في الصين مثلًا، أكثر من ١٢٠٠ فرع لـ ستاربكس، وتقوم الشركة بافتتاح فرع جديد كل يوم تقريبًا. وفي سبيل تحديد موقع الفرع التالي

تكــون الشركة بحاجة لكثير من المعلومات عن المناطق التجارية وتجمعات بيع التجزئة والمولــدات والحركة المروريــة والمعلومات الديموغرافية. وبعد تحليل الســوق والحي الجديدين، يســتطيع المدير الحصول على نظرة مكــبرة لمواقع محددة في أي منطقة في المدينة، ومن ثَمَّ تحديد ثلاثة أبراج مكتبية جديدة يمكن أن تكتمل في غضون شــهرين مثلًا. وبعد عرض هذه المنطقة على الخريطة، يمكن إنشــاء نافذة لســير العمل والتي ســوف تساعد المدير على متابعة حركة الموقع الجديد بدءًا بالموافقة، مرورًا بالتصاريح، ثم التشييد، وأخيرًا الافتتاح.

ويمكن أيضًا إدارة عمليات الطلب وسلاســل التوريد بشكل أفضل، وذلك من خلال تكامل بيانات الطقس مع البيانات المحلية الأخرى. وتســعى ستاربكس لتكامل أنظمة أعمالها مع حلول نظم (GIS) الخاصة بها في خدمات الويب للحصول على رؤية جديدة للأعمال في جميع أنحاء العالم. فمثلًا، تقوم ستاربكس بدمج توقع بيانات درجة الحرارة الحقيقيــة التي تتنبــأ بها أرصاد. ويمكن أن تســاعد هذه البيانــات المتوقعة لدرجات الحرارة في تمركز الجهود التســويقية. ومع اقتراب الأســبوع شديد الحرارة في ممفيس، يمكن لمحللي ســتاربكس اختيــار مجموعة من بيوت القهــوة والحصول على معلومات مفصلــة عن أنماط الطقس في الماضي والمســتقبل، بالإضافة إلى خصائص الفرع. ويمكن اســتخدام هذه المعرفة لتصميم ترويج محلي لفرابتشينو Frappuccinos. فمثلًا، يمكن مساعدة ستاربكس على توقع ما سيطلبه عملاؤه لأسبوع مقدمًا.

وللأحــداث الكبيرة أيضًــا أثرٌ على بيوت القهوة. فعندما نزل ١٥٠ ألف شــخص في موكب كبير في San Diego، قام الســقاة المحليون بخدمة الكثير من العملاء. ولضمان أفضــل تجربة ممكنة للعميل، اســتخدمت ســتاربكس هذه المعرفــة بالحدث المحلي لتخطيط طاقم العمل والمخزون في مواقع قريبة من الموكب.

أسئلة للمناقشة:

١- ما هو نوع المعلومات الديموغرافية ومعلومات GIS التي قد تتعلق بقرار تحديد موقع الفرع؟

٢- من المعلوم أن ســتاربكس تشجع عملاءها على استخدام تطبيق الجوال الخاص بها. ما هو نوع المعلومات التي ربما تلتقطها الشركة من التطبيق لمساعدتها في التخطيط الأفضل للعمليات؟

٣- هل إتاحة Wi-Fi بالمجان في فروع ستاربكس ستقدم أي معلومات إلى ستاربكس من أجل تحليلات أفضل؟

Sources: Digit.HBS.org (2015). Starbucks: Brewing up a data storm! https://digit.hbs.org/submission/Starbucks-brewing-up-a-data-storm/(accessed August 2016); Wheeler, C. (2014). Going Big with GIS. http://www.esri.com/esri-news/arcwatch/0814/going-Big-with-gis (accessed August 2016); Blogs.ESRI.com. From customers to CxOs, Starbucks delivers world-class service. (2014). https://blogs.esri.com/esri/ucinsider/201429/07//Starbucks/(accessed August 2016).

وبالإضافــة إلى تطبيقات تحليل معاملات التجزئة التي أبرزناهــا هنا، فهناك تطبيقات أخرى متعددة للجمع بين المعلومات الجغرافية وبين البيانات الأخرى التي يتم إنشاؤها من قِبل المنظمة. فعلى ســبيل المثال، غالبًا ما تنتج عمليات الشــبكة وشركات الاتصالات كميات هائلة من البيانات اليوم. وتسطيع القدرة على تحليل البيانات بسرعة مع مستوى عالٍ من التفاصيل الخاصة بالموقع أن تحدد بشــكل أفضل الاضطراب عند العميل ومن ثّم تقوم بالمســاعدة في صياغة إستراتيجيات خاصة بالمواقع من أجل زيادة كل من الكفاءة التشغيلية، وجودة الخدمة، والأرباح.

ويســتطيع التحليل الجغرافي المكاني أن يُمَكِّن شركات الاتصالات من رصد المعاملات اليومية من الشــبكة لتحديد المناطق الجغرافية التي تعاني من فشــل عدد كبير من محاولات الاتصال الخاصة بالصوت، أو البيانات، أو النص، أو الإنترنت. وتســتطيع التحليلات أن تســاعد في تحديد الأســباب الدقيقــة اعتمادًا على الموقع والانتقال إلى عميل فردي من أجل توفير خدمة عملاء أفضل. ويمكنك أن تلمس ذلك أثناء العمل عن طريق إكمال تمرين الوسائط المتعددة التالي.

تمرين الوسائط المتعددة في تحليلات تستخدم التحليلات الجغرافية المكانية:

تشتمل شبكة جامعة تيراداتا على فيديو BSI في حالة فشل مكالمات الهاتف المحمول. يُرجى مشاهدة الفيديو المتاح على موقع يوتيوب عبر الرابــط التالي: http://www.teradatauniversitynetwork.com/Library/Items/BSI-The-Case-of-the-Dropped-Mobile-Calls./

وتطلق شركة اتصالات خطًا جديدًا من الهواتف الذكية وتواجه مشاكل مع انخفاض المكالمات. وتتعرض عملية الطرح الجديد إلى مشــاكل، وتعتبر المنطقة الشــمالية الشرقية هي المنطقة الأكثر تضررًا حيــث يقومون بمقارنة تأثيرات المكالمات المفقودة على أربــاح المنطقة الجغرافية. وتقوم الشركة بإيجار BSI لتحليل المشــاكل الناجمة عن عيوب الهواتف الذكية، وتغطية البرج، ومواطن

خلل البرمجيات. وتنقسم بيانات المنطقة الشمالية الشرقية بأكملها إلى مجموعات جغرافية، وتقوم الشركة بحل المشكلة من خلال تحديد بيانات العملاء الفردية. ويستخدم فريق BSI التحليلات الجغرافية المكانية لتحديد المواقع التي تؤدي فيها تغطية الشبكة إلى انخفاض المكالمات وتقترح تثبيت بضعة أبراج إضافية في موقع العملاء غير الراضين عن الخدمة.

وبعد اكتمال الفيديو، يمكنك الاطلاع على كيفية إعداد التحليل عبر:

slideshare.net/Teradata/bsi-Teradata-the-case-of-the-dropped-mobile-calls.

وتقدم هذه النزهة من الوسائل المتعددة مثالًا على مجموعة من التحليلات الجغرافية المكانية بجانب تحليلات البيانات الضخمة التي تساعد في اتخاذ القرارات بشكل أفضل.

ذكاء الموقع الفوري:

تقوم العديد من الأجهزة المستخدمة بواسطة المستهلكين والمهنيين بإرسال معلومات الموقع الخاصة بهم بوتيرة ثابتة. وتقوم السيارات والحافلات وسيارات الأجرة وأجهزة الجوال والكاميرات وأجهزة التنقل الشخصية ببث إشارات لمواقعها باستخدام تقنيات تحديد المواقع المتصلة بالشبكة مثل (GPS) وWi-Fi وcell tower triangulation.

ويستخدم الملايين من المستهلكين وأصحاب الأعمال الأجهزة الممكّنة بالموقع للعثور على خدمات قريبة، وموقع الأصدقاء والعائلة، والتنقل، وتعقب الأصول والحيوانات الأليفة، والاشتراك في الألعاب الرياضية، والهوايات. وقد أدى هذا الاندفاع القوي في الخدمات الممكّنة بالموقع إلى قاعدة بيانات ضخمة من المعلومات التاريخية ومعلومات الموقع المتدفقة أولًا بأول، وهي بالتأكيد متفرقة وغير مفيدة في حد ذاتها. حيث تقدم مجموعة البيانات الآلية التي تم تمكينها من خلال التقاط الهواتف الخلوية ونقاط الوصول الساخنة إلى Wi-Fi بُعدًا جديدًا مثيرًا للاهتمام في أبحاث السوق غير المتداخلة، وجمع البيانات، وبالطبع، التحليل الدقيق لمثل هذه المجموعات الضخمة من البيانات.

ويمكن من خلال التحليل والتعلم من هذه الأنماط واسعة النطاق للحركة، أن يتم تحديد فئات متميزة من السلوكيات في سياقات محددة. ويسمح هذا النهج لأصحاب الأعمال بفهم أنماط عملائهم بشكل أفضل واتخاذ قرارات أقوى بشأن العروض الترويجية والتسعير وما إلى ذلك. كما يمكن من خلال تطبيق الخوارزميات التي تقلل من أبعاد بيانات الموقع، أن يتم تمييز الأماكن وفقًا للنشاط والتنقل بينها. ومن الكميات الهائلة لبيانات الموقع ذات الأبعاد المرتفعة، تكشف هذه الخوارزميات عن اتجاهات ومعانٍ وعلاقات تنتج في النهاية تمثيلات سهلة الفهم للبشر، ثم يصبح

بعد ذلك من الممكن استخدام مثل هذه البيانات لعمل تنبؤات ذكية تلقائيًا وإيجاد أوجه تطابق وتشابه مهمة بين الأماكن والأفراد. ويظهر تطبيق التحليلات القائمة على الموقع في تطبيقات التسويق الموجهة بالمستهلك. وقد قام Quiznos، وهومطعم خدمات سريعة، باستخدام منصة Sense Networks لتحليل مسارات المواقع لمستخدمي الجوال بناءً على البيانات الجغرافية المكانية المتحصل عليها من نظام (GPS) والعملاء المستهدفين من الذكاء التقني باستخدام الكوبونات. انظر حالة عملية (٨-٦). توضح هذه الحالة الاتجاه الحديث في مجال بيع التجزئة حيث تتطلع الشركات إلى رفع كفاءة الحملات التسويقية، ليس فقط من خلال استهداف كل عميل بناءً على موقعه الحالي، ولكن عن طريق استخدام تحليلات تنبؤية أكثر تطورًا في الوقت المناسب على خصائص المستهلكين السلوكية لإيجاد المجموعة الصحيحة من المستهلكين للحملات الإعلانية.

حالة عملية ٨-٦

مطعم Quiznos يستهدف عملاء للسندويشات التي يقدمها

قام مطعم Quiznos الحاصل على حق امتياز للخدمة السريعة، بتنفيذ حملة استهداف أجهزة الجوال القائمة على الموقع والتي استهدفت مستهلكين ذوي دهاء تقني وغير متفرغين من مدينة بورتلاند، بولاية أوريجون. واستخدمت الحملة منصة Sense Networks، والتي قامت بتحليل مسارات مواقع مستخدمي الجوال لفترات زمنية مفصّلة، وقامت بإنشاء ملفات شخصية مجهولة بناءً على السمات السلوكية لعادات التسوق.

وبتطبيق التحليلات التنبؤية على الملفات الشخصية للمستخدمين، قام مطعم Quiznos باستخدام الاستهداف السلوكي المبني على الموقع لتقليل خصائص المستخدمين المحتمل تناولهم للأكل في مطعم الخدمة السريعة أكثر من غيرهم. واستمرت هذه الحملة الإعلانية لمدة شهرين - نوفمبر وديسمبر ٢٠١٢ - واستهدفت فقط العملاء الذين تتراوح أعمارهم بين ١٨ و٣٤ عامًا والمحتمل تواجدهم في مطاعم الخدمة السريعة خلال الثلاثين يومًا السابقة، في حدود ٣٠ ميلًا مربعًا تقريبًا حول مطعم Quiznos. واستخدمت إعلانات الجوال ذات الصلة بالكوبونات المحلية بناءً على موقع العميل. وقد نتج عن هذه الحملة أكثر من ٣,٧ مليون عميل جديد وزيادة بنسبة ٢٠٪ في عمليات استرداد الكوبونات داخل منطقة بورتلاند.

أسئلة للمناقشة:

١- كيف يمكن أن تساعد التحليلات القائمة على الموقع تجار التجزئة في استهداف العملاء؟

٢- قم بالبحث عن تطبيقات مشابهة للتحليلات قائمة على الموقع في مجال بيع التجزئة.

Source: Adapted from Mobilemarketer.com (2013). Quiznos sees 20pc boost in coupon redemption via location-based mobile ad campaign. mobilemarketer.com/cms/news/advertising/14738.html (accessed August 2016).

ولا زال هناك امتداد آخر للتحليلات القائمة على الموقع وهو استخدام الواقع المعزز. في عام ٢٠١٦، ظهرت ضجة سوقية اسمها Pokémon GO، وهي لعبة قائمة على الواقع معززة باستشعار الموقع حيث تشجع المستخدمين على المطالبة بأشياء افتراضية من المواقع الجغرافية المنتقاة. فيستطيع المستخدم أن يبدأ من أي مكان في المدينة ويسير وفقًا لعلامات على التطبيق للوصول إلى بند معين. وتكون الأشياء الافتراضية مرئية من خلال التطبيق عندما يوجه المستخدم كاميرا الهاتف باتجاه أي منها، ويستطيع المستخدم بعد ذلك أن يطالب بذلك الشيء. وتعد تطبيقات الأعمال لمثل هذه التقنيات أيضًا حديثة الظهور. على سبيل المثال، يوجد تطبيق يسمى Candybar، يسمح هذا التطبيق لأصحاب الأعمال بوضع هذه الأشياء الافتراضية على خريطة باستخدام خرائط Google. كما يمكن باستخدام Google Street View تحديد مواقع تلك الأشياء بمزيد من الدقة، وبمجرد أن يتم تنسيق كل البنود الافتراضية مع المعلومات والموقع، يصبح بإمكان صاحب العمل أن يبث بنودًا مرئية للمستخدم بشكل فوري. كما يوفر Candybar أيضًا تحليلات استخدام الأعمال لتمكين استهداف أفضل للأشياء الافتراضية. ويعمل جانب الواقع الافتراضي لهذا التطبيق على تحسين تجربة المستخدمين، بتوفير بيئة "ألعاب" لهم في الحياة الواقعية. وفي نفس الوقت، يوفر منصة تسويق قوية لأصحاب الأعمال للوصول إلى عملائهم.

وكما هو واضح من هذا القسم، فقد تصبح التحليلات القائمة على الموقع والتطبيقات الناتجة عنها هي أهم واجهة للمنظمات في المستقبل القريب. وكان استخدام البيانات التشغيلية أو التسويقية من قِبل المنظمات واحدًا من الأفكار الشائعة في هذا القسم. وسوف نقوم لاحقًا باستكشاف تطبيقات التحليلات التي تستهدف المستخدمين بشكل مباشر وفي بعض الأحيان تستغل ميزة معلومات الموقع.

تطبيقات التحليلات للمستهلكين:

لقد أوجد النمو الهائل في صناعة التطبيقات لمنصات الهواتف الذكية (iOS وAndroid وWindows وما تبعها) وفي استخدام التحليلات فرصًا كبيرة لتطوير التطبيقات حيث يستخدم المستهلكون التحليلات دون إدراك. وتختلف هذه التطبيقات عن الفئة السابقة في أنها معنية

بالاستخدام المباشر للمستهلك، بدلًا من المنظمة التي تحاول التنقيب في بيانات المستخدم المتعلقة بالاستخدام / بالشراء لإنشاء ملف تعريف لتسويق منتجات أو خدمات محددة. ووفقًا للتنبؤات، تعمد هذه التطبيقات إلى تمكين المستهلكين من اتخاذ قرارات أفضل عن طريق استخدام تحليلات محددة. ونستعرض فيما يلي مثالين على ذلك.

يعمل Waze - وهو تطبيق ويب اجتماعي - على مساعدة المستخدمين في تحديد مسار التنقل وتنبيه المستخدمين بالمشكلات المحتملة مثل الحوادث ونقاط التفتيش وأجهزة مراقبة السرعة ومواقع أعمال البناء على الطريق، وذلك بناءً على مدخلات مستخدمين آخرين. وقد أصبح Waze أحد أكثر تطبيقات التنقل انتشارًا. وقد حصلت Google على هذا التطبيق منذ بضع سنوات، وعملت على تحسينه. ويعد هذا التطبيق مثالًا على تجميع المعلومات التي ينشئها المستخدم وجعلها متاحة للعملاء.

تسمح الكثير من التطبيقات للمستخدمين بإرسال الاستعراضات والتصنيفات لأصحاب الأعمال والمنتجات وما إلى ذلك، ومن ثَمّ تقديمها للمستخدمين في صورة متكاملة لمساعدتهم في وضع خيارات. ويمكن تعريف هذه التطبيقات أيضًا كتطبيقات مبنية على بيانات اجتماعية تستهدف المستهلكين حيث يقومون بإنشاء البيانات. ويعتبر Yelp أحد التطبيقات الأكثر رواجًا بين تطبيقات الفئة.

وتم نشر تطبيق آخر مرتبط بالنقل يستخدم التحليلات التنبؤية في بيتسبيرج، بنسلفانيا. وقد تم تطويره بالتعاون مع جامعة Carnegie Mellon، ويشمل هذا التطبيق إمكانيات تنبؤية لتقدير توافر مواقف السيارات. ويقوم ParkPGH بتوجيه السائقين إلى المناطق التي تتوفر فيها مواقف للسيارات. فهو يقوم بحساب عدد أماكن وقوف السيارات المتاحة في ١٠ مواقف والتي تتجاوز ٥٣٠٠ مساحة و٢٥٪ من مواقف السيارات في وسط مدينة بيتسبيرج. ويتم تحديث المساحات المتاحة كل ٣٠ ثانية، مما يحفظ تواجد السائق بالقرب من الأماكن المتاحة قدر الإمكان. وبالاعتماد على الطلب التاريخي والأحداث الحالية، يكون التطبيق قادرًا على التنبؤ بتوفر مواقف للسيارات وتوفير معلومات عن المساحة التي ستكون متاحة مع الوقت الذي يصل فيه السائق إلى المكان المقصود. وتستخدم الخوارزمية الأساسية للتطبيق بيانات حول الأحداث الجارية حول المنطقة - على سبيل المثال، لعبة كرة السلة - للتنبؤ بزيادة في الطلب على أماكن وقوف السيارات في وقت لاحق من ذلك اليوم، وبالتالي توفير وقت ثمين للمسافرين للبحث عن أماكن وقوف السيارات في المدينة المشغولة.

لم تكن نشأة التطبيقات القائمة على التحليلات فقط للمرح والصحة، ولكنها ظهرت كذلك لتعزيز إنتاجية الفرد. فعلى سبيل المثال، تتواجد التطبيقات Cloze وCRM وغيرها في مكان

واحـد. ويقـوم تطبيـق Cloze بإدارة صناديق الرسـائل الـواردة للعديد من حسـابات البريد الإلكتروني بجانب حسـابات أخرى لوسـائل التواصـل الاجتماعي. فهو يقـوم بعمل تكامل بين الشـبكات الاجتماعية وجهات اتصال البريد الإلكتروني لمعرفـة أي من جهات الاتصال تكون أكثر أهميـة من غيرها ويقوم بإعطائها درجات، بحيث تحصل جهات الاتصال المهمة علي درجة أعلى. فيتم عرض رسـائل البريد الإلكتروني ذات الدرجات الأعـلى أولًا، وبالتالي يقوم بإبعاد البريد الأقل أهمية وغير ذي الصلة. ويعمل Cloze على تخزين سياق كل محادثة لتوفير الوقت عند العودة لمكالمة معلقة. ويتم تنظيم جهات الاتصال في مجموعات بناءً على عدد مرات تكرار تفاعلها، مما يسـاعد المسـتخدمين على البقاء على اتصال مع الأفراد الذين قد يفقدون الاتصال بهم. ويمكن للمسـتخدمين وضع درجات Cloze للأفراد الذين يريدون الحصول على اتصال بهم والعمل على تحسين هذه الدرجة. ويقوم Cloze برفع درجة كلما حدثت محاولة اتصال. فعلى سبيل المثال، عنـد فتح أي بريد إلكتروني يوفـر Cloze عدة خيارات، مثل "الآن"، "اليوم"، "غدًا"، "الأسـبوع القادم"، والتي تُذكِّر المستخدم تلقائيًا أن يبدأ الاتصال وفقًا للجدول الزمني. ويعمل هذا كتذكير للرجوع إلى رسائل البريد الإلكتروني في وقت لاحق، دون نسيانها أو وضع علامة "غير مقروء"عليها، والتي غالبًا ما تؤدي إلى ازدحام صندوق الوارد. ولأن Cloze الآن أصبح مستهدفًا كتطبيق إنتاجي للأعمال، تعتبر أسعاره حاليًا أكبر من إمكانيات المستهلك.

وكما يتضح من هذه الأمثلة للتطبيقات المرتكزة على المسـتهلك، تبدأ التحليلات التنبؤية في تمكين تطوير البرمجيات التي يسـتخدمها المسـتهلك مباشرة. وتقدِّر مجلة Wall Street (wsj.com/apps) حجـم صناعة التطبيقات الذي وصل بالفعل لقيمة ٢٥ مليار دولار بنمو متوقع أكبر. ونحن نؤمن بأن نمو التطبيقات التحليلية الموجهة بالمسـتهلك سيستمر ويصنع الكثير من فرص الأعمال الحرّة لقراء هذا الكتاب.

ويتمثل أحد أهم المخاوف من اسـتخدام هذه التقنيات في فقدان الخصوصية. فإذا اسـتطاع أي شخص تتبع حركة انتقال الهاتف الخلوي، تكون خصوصية هذا العميل مشكلة كبيرة؛ إذ يدعي بعـض مطوري التطبيقات أنهم بحاجة فقط إلى جمع معلومات تدفق مجمعة، وليس المعلومات سهلة التحديد بشكل فردي، في حين تظهر الكثير من القصص في وسائل الإعلام تسلط الضوء على اختراقات لهذا المبدأ العام. ويجب أن يكون لدى كل من المستخدمين والمطورين لهذه التطبيقات الوعـي التام بالأثـر الضار الناجم عن منح معلومات خاصة فضلًا عن جمع مثل هذه المعلومات. وفي القسم التالي سوف نناقش هذه القضية بشيء من التفصيل.

أسئلة مراجعة على القسم ٨-٤:

١- كيف تستخدم التحليلات التقليدية البيانات القائمة على الموقع؟

٢- كيف يمكن أن تساعد المواقع المكودة جغرافيًّا في اتخاذ قرار أفضل؟

٣- ما هي القيمة المضافة بواسطة التحليلات الجغرافية المكانية؟

٤- قـم باستكشـاف اسـتخدام التحليـلات الجغرافية المكانية بشـكل أكبر من خـلال التحقق من استخدامها عبر قطاعات متنوعة مثل تتبع التعداد الحكومي، والتسويق للمستهلك، وما إلى ذلك.

٥- قم بالبحث على الإنترنت عن تطبيقات أخرى للتطبيقات التحليلية الموجهة بالمستهلك.

٦- كيف يمكن للتحليلات القائمة على الموقع أن تساعد المستهلكين الأفراد؟

٧- قم باستكشاف المزيد من تطبيقات النقل التي قد تستخدم تحليلات قائمة على الموقع.

٨- ما هي التطبيقات الأخرى التي يمكنك تخيلها إذا كنت قادرًا على الوصول إلى بيانات موقع الهاتف الخلوي؟

٨-٥ القضايا القانونية، والخصوصية، والأخلاقية:

لقـد تطورت تطبيقات علم البيانات، والتحليلات، والحوسـبة الإدراكيـة، بالإضافة إلى الذكاء الاصطناعي في إمكانية الوصول إليها وانتشـارها، وقد أدى هذا التطور إلى تأثر كل شـخص بهذه التطبيقات. ومجرد إمكانية تحقيق شيء من خلال التقنية، لا تجعله شـيئًا مناسـبًا، أو قانونيًّا، أو أخلاقيًّا. ويجب أن يكون متخصصو علم البيانات والمديرون على وعي تام بهذه المخاوف. وترتبط العديد من القضايا القانونية، والخصوصية، والأخلاقية المهمة بالتحليلات. ونقدم هنا أمثلة معبرة ومصادر فقط. وعادةً ما تكون وسـائل الإعلام المنتشرة حريصة للغاية على الإبلاغ عن مثل هذه الانتهاكات للسـلوك القانوني والأخلاقي. لذلك؛ فإن هذا القسـم هو أحد الأقسام الذي قد يمكنك العثـور فيه على المزيد من الأمثلة الحديثة عبر الإنترنت. وكما أشرنا في الفصل الأول، فإننا نهدف هنـا فقـط لإطْلاعك على هذه القضايـا. وينبغي على فِرَق الفصل الخاص بـك أن تحدد حالات أحدث وتقوم بمناقشة هذه الحالات.

القضايا القانونية:

قد يؤدي إدخال التحليلات إلى تفاقم مجموعة كبيرة من القضايا القانونية التي لها صلة فعلية بأنظمة الحاسب. وعلى سبيل المثال، فإن الأسئلة المتعلقة بالمسؤولية عن أعمال المشورة المقدمة من الآلات الذكية بدأت تؤخذ بعين الاعتبار.

وبالإضافــة إلى حل النزاعات حول نتائج بعض التحليــلات غير المتوقعة بل وربما الضارة، فإنه قد تظهر قضايا أخرى معقدة، مثل: مَن الذي يتحمل المســؤولية في حالة ما إذا وجدت المؤسسة نفســها مفلسة نتيجة لاستخدام نصيحة تطبيق تحليلي؟ هل ستتحمل المؤسسة المسؤولية لعدم اختبار النظام بشكلٍ كافٍ قبل إسناد قضايا حساسة له؟ هل ستشارك شركات التدقيق والمحاسبة في المســؤولية عن الفشــل في تطبيق اختبارات تدقيق كافية؟ هل ســيكون لمطوري البرمجيات الخاصة بالأنظمة الذكية دور في المســؤولية؟ عندما تصبح الســيارات ذاتية القيادة أكثر انتشارًا، من ســيكون المسؤول عن أي ضرر أو تلف عندما تتعطل أجهزة استشعار السيارة، أو الشبكة، أو التحليلات في العمل كما هو مُخطط لها؟ وهناك حالة حديثة تتعلق بحادث سيارة Tesla حيث مات السائق بينما أظهرت السيارة أنها في وضع الطيار الآلي «autopilot» وقد نُقلت هذه القضية على الصفحات الأولى من الصُحُف والمهن القانونية.

وقدمت الحالة العملية (٧-١) أمثلة على اســتخدام البيانات البديلة في اســتخدام معلومات إضافية من أجل توقعات صائبة بشــأن مخرجات الحصيلة المستقبلية، ومبيعات الشركة، وغيرها. وعــلى الرغم من أن هــذه الأدوات اليوم لا تفرض قيودًا تقليدية على التــداول من الداخل نظرًا لعدم مشاركة أي شــخص داخل الشركة للمعلومات غير العامة، إلا أنه ربما لا تزال هناك مشاكل حول الحصول على المعلومات التي تكون غير متاحة للعامة. وقد لاحظ Ekster (2015) أن أي تاجر في الســوق يستخدم التحليلات والبيانات البديلة لابد أن يكون على علم بقواعد المعلومات المفروضــة مــن قِبل الهيئات التنظيمية الماليــة. وعادةً ما تكون الأفكار المســتمدة من البيانات العامة آمنة نظرًا لعدم الحصول عليها بشكل مباشر من مصادر داخلية.

تأمل المسائل القانونية المُحددة التالية:

- ما قيمة رأي الخبير في المجال إذا ما تم تشفير الخبرة في الحاسب؟
- من هو المســؤول عن النصيحة (أو المعلومات) الخاطئة التــي يقدمها أحد التطبيقات الذكية؟ فمثلًا، ماذا يحدث إذا صدق الطبيب تشخيصًا غير صحيح شخَّصه جهاز الحاسب وقام الطبيب على أساسه باتخاذ إجراء تسبب في وفاة أحد المرضى؟
- مــاذا يحــدث لو قام أحد المديرين بإدخال تقييم حكمي غــير صحيح في تطبيق تحليلي وكانت النتيجة حدوث ضرر أو كارثة؟
- من الذي يملك المعرفة في قاعدة المعرفة؟
- هل تملك الإدارة الحق في إجبار الخبراء على المساهمة بخبراتهم؟

الخصوصية:

يختلف مفهوم الخصوصية باختلاف الناس. وبشكل عام، فإن الخصوصية هي حقك في أن تنفرد بذاتــك، كما أنها الحق في أن تتحرر من تدخلات الآخرين غير المعقولة في أمورك الشــخصية. ولطالما كانــت الخصوصية قضية قانونيــة، وأخلاقية، واجتماعية في العديد من البلــدان، حيث يتم الاعتراف بالحــق في الخصوصية اليوم في كل ولاية من الولايات المتحدة الأمريكية والحكومة الفيدرالية، ســواء أكان من خلال التشريع، أو بموجب القانون العام. ويمكن تفســير تعريف الخصوصية تفســيرًا واسعًا جدًا، بيد أن المحكمة كانت تتبع في قراراتها السابقة القاعدتين التاليتين بشكل دقيق إلى حد ما، وهما:

١- أن حق الخصوصية ليس حقًا مطلقًا. فالخصوصية يجب أن تكون متوازنة مع احتياجات المجتمع.

٢- أن حق الجمهور في المعرفة يفوق حق الفرد في الخصوصية.

وتوضح هاتان القاعدتان ســبب صعوبة تحديــد وتطبيق قوانين الخصوصية في بعض الأحيان (انظــر Peslak, 2005). وتتضمــن قضايا الخصوصية عبر الإنترنت خصائص وسياســات محددة. وفيما يلي ســنناقش أحد المجالات الذي قد تتعرض فيه الخصوصية للخطر. وللتعرف على قضايا الخصوصية والأمان في بيئة مســتودع البيانات، يمكنك أيضًا الاطلاع على الورقة التي كتبها كل من Elson وLeClerc (2005).

جمع المعلومات عن الأفراد:

في كثــير من الحالات كان التعقيد في عملية جمع المعلومات، وفرزها، وحفظها، والوصول إليها يدويًــا في العديد مــن الوكالات الحكومية، بمثابة وقاية تلقائية ضد إســاءة اســتخدام المعلومات الخاصة. فقد كان الإقدام على انتهاك خصوصية أي شخص ببساطة أمرًا مكلفًا، ومرهقًا، فضلًا عن كونه معقدًا. وقد أنشأ الإنترنت بعدًا جديدًا تمامًا للوصول إلى البيانات واستخدامها، جنبًا إلى جنب مع قواعد البيانات واسعة النطاق. ويمكن استخدام القوة الكامنة في الأنظمة التي يمكنها الوصول إلى كميات هائلة من البيانات لأجل مصلحة المجتمع. فمثلًا، من خلال مطابقة السجلات بمساعدة جهاز الحاســب، يمكن القضاء على أو الحد من الاحتيال والجريمة وسوء الإدارة الحكومية والتهرب الضريبــي والغش في الرعاية الاجتماعية وسرقة الدعم الأسري وتوظيف العمال غير الشرعيين، وغير ذلــك. ولكــن ما هو القدر الذي يجــب أن يتكلفه الفرد من فقدان الخصوصية بحيث تســتطيع الحكومــة إلقاء القبض على المجرمين بشــكل أفضل؟ وينطبق الأمر ذاته على مســتوى الشركات. فعلى الرغم من أن المعلومات الخاصة بالموظفين قد تساعد في اتخاذ القرارات بشكل أفضل، إلا أن خصوصية الموظفين قد تتأثر بهذا الأمر. وهناك قضايا مشابهة تتعلق بالمعلومات الخاصة بالعملاء.

الجدير بالذكر أن الآثار المترتبة على الخصوصية على شـبكة الإنترنت لا يُسـتهان بها. ويوسع القانون الوطني الأمريكي أيضًا من قدرة الحكومة على الوصول إلى معلومات الطالب والمعلومات المالية الشـخصية دون أي شـبهة في ارتكاب مخالفات، وذلك عبر إثبات أن تلك المعلومات التي مـن المُرجح الحصول عليها لها صلة بالتحقيق الجنـائي الجاري (انظر مركز معلومات الخصوصية الإلكترونية، ٢٠٠٥). فقد تم استخدام معلومات الموقع من الأجهزة لتحديد مكان الضحايا وكذلك الجنـاة في بعـض الحـالات، ولكن إلى أي حد تكون المعلومات ليسـت ملكًا للفـرد؟ وقد جلب اللغـط الأخير حول ما تقـوم به الولايات المتحدة وغيرها من الدول الأخرى من تسـجيل بيانات المكالمـات الهاتفية، والبريد الإلكتروني، وغيرها من التحركات المرورية الإلكترونية العديد من هذه القضايـا إلى المقدمة. وبالإضافة إلى ذلك، فقـد أدى تسريب المعلومات من الاتصالات الحكومية عـن طريق Wikileaks إلى حدوث عاصفة كبيرة على مدى السـنوات القليلة الماضية. وقد أدى إفصـاح Edward Snowden مؤخـرًا، عن برامج الحكومة الأمريكية لجمع البيانات واسـتخراج هـذه المعلومات أيضًا إلى ضجة إعلامية كبيرة حول قضايا انتهاك الخصوصية. وأي بحث في هذه الموضوعات سوف يسفر عن الكثير من الروابط ووجهات النظر. وبصفتك محترفًا لعلم البيانات، فإنه ينبغي عليك أن تدرك أن مثل هذه القضايا يمكنها أن تُحدث فرقًا كبيرًا في سمعة منظمتك.

وهنـاك أداتان فعالتان لجمـع المعلومات حول الأفراد وهما ملفات تعريـف الارتباط وبرامج التجسس. وقد بدأت وسائل تسجيل الدخول الأحادي التي تسمح للمستخدم بالوصول إلى خدمات متنوعة من مُقَدِم الخدمة تثير بعض نفس المخاوف مثل ملفات تعريف الارتباط. وهناك خدمات مثـل (Google، وYahoo، وMSN) تطلـب من المسـتهلكين بشـكل دائم إدخـال ملف تعريف المعلومـات مع كلمة مرور واسـتخدام هذه المعلومات وكلمة المرور بشـكل متكـرر للوصول إلى الخدمـات في مواقـع متعددة. وقد ذكر النقاد أن مثل هذه الخدمات ينتج عنها نفس الفرص التي تنتج عن ملفات تعريف الارتباط لانتهاك خصوصية الفرد.

وقد يؤدي اسـتخدام تقنيات الـذكاء الاصطناعي في الإدارة وتطبيق القوانـين واللوائح إلى زيادة المخاوف العامة فيما يتعلق بخصوصية المعلومات. وهذه المخاوف، التي نتجت عن القدرات الملموسة للذكاء الاصطناعي، سوف تتعين معالجتها في بداية أي محاولة لتطوير الذكاء الاصطناعي تقريبًا.

خصوصية مُستَخدِم الهاتف المحمول:

هناك العديد من المُسـتَخدِمين غير مُدركِين للمعلومات الشخصية التي يتم تعقبها من خلال اسـتخدامهم للهواتف الذكية. وهناك العديد من التطبيقات التي تجمع بيانات المُسـتَخدِم من خـلال تعقب كل هاتف أثنـاء انتقاله من بُرج خلية إلى آخر، ومن خـلال أجهزة تحديد المواقع

والتي تنقل مواقع المُستَخدمين، فضلًا عن تجميع البيانات من الهواتف التي تنقل المعلومات عبر نقـاط اتصـال Wi-Fi. وعلى الرغم من زعم مطوري التطبيقات الأساسـيين أنهم دقيقون للغاية ويعملون على حماية خصوصية المستخدمين، إلا أنه من المثير للاهتمام ملاحظة مقدار المعلومات المتوفـرة مـن خلال اسـتخدام جهاز واحد. وقد حـدث جدل مؤخرًا بين كل مـن شركة Apple والحكومـة الأمريكية حول طلب حكومة الولايات المتحدة لإلغاء قفل الأيفون (iPhone) ورفض شركة آبل القيام بتعديل برامجها من أجل تلبية هذا الطلب. حيث تزعم شركة Apple أنها تجمع معلومات قليلة أو تكاد تكون معدومة حول عملائها من مُستخدمي أيفون. وعلى الجانب الآخر، فـإن Google يجمـع الكثير من المعلومات المفيدة حتى يتمكن من تقديم معلومات اسـتباقية للمُسـتَخدمين حـول التقويم الخاص بهـم، والمطاعم المُفَضلة، والأنشـطة القادمة، وغيرها. كما ينبغي على أي مُطور تطبيق تحليلات أن يضع في اعتباره قضايا الخصوصية.

الأمن الداخلي والخصوصية الفردية:

على الرغم من أن الكثيرين يعتبرون أن اسـتخدام تقنيات التحليلات مثل اسـتخراج وتفسـير محتـوى المكالمـات الهاتفية، والتقـاط صور للأشـخاص في أماكن معينة وتحديدها، واسـتخدام الماسحات الضوئية لرؤية متعلقاتك الشخصية، بمثابة انتهاك للخصوصية، إلا أن العديد من الناس يدركون أن تلك الأدوات التحليلية هي وسـيلة فعالـة وذات كفاءة لزيادة الأمن، على الرغم من تعرُّض خصوصية العديد من الأبرياء للانتهاك.

وتقـوم حكومة الولايات المتحـدة بتطبيق تقنيات تحليلية على نطـاق عالمي في الحرب على الإرهاب. ففي العام ونصف الأول بعد أحداث الحادي عشر من سـبتمبر ٢٠٠١، قامت سلاسـل محلات السـوبر ماركت، ومتاجر مسـتلزمات تجديد المنازل، وغيرهم من تجار التجزئة بتسـليم كميات هائلة من سـجلات العملاء طوعًا إلى وكالات تنفيذ القانون الفيدرالية، منتهكة في الغالب سياسـات الخصوصية المُعْلَنَة. وقد اسـتجاب الكثير من الآخرين لأوامر المحكمة في الحصول على معلومـات، كما يقتـضي القانون. وتملك حكومة الولايات المتحـدة الحق في جمع بيانات الشركة بموجـب التشريع الذي تم تمريـره بعد أحداث الحادي عشر من سـبتمبر ٢٠٠١. ويقوم مكتب التحقيقات الفيدرالي (FBI) الآن باسـتخراج كميات هائلة من البيانات، بحثًا عن أي نشاط يمكن أن يُشير إلى وجود مؤامرة إرهابية أو جريمة.

الجدير بالذكر أن قضايا الخصوصية كثيرة. ونظرًا لأن الحكومة تحصل على البيانات الشخصية للكشف عن أنماط النشاطات المريبة، فهناك احتمال أن تُستَخدم هذه البيانات بشكل غير لائق أو

غير قانوني. ويرى الكثيرون أن مجرد جمع البيانات يُعد انتهاكًا لحريات وحقوق المواطنين. فهم يــرون ضرورة وجــود منظمة رقابة «للإشراف على المراقبين»، للتأكــد من أن وزارة الأمن الداخلي لا تحصــل عــلى البيانــات دون داعٍ لذلك. وبدلًا من ذلــك، فإنه ينبغي الحصــول على البيانات والمعلومات التي لها صلة وثيقة بالموضوع فقط والتي يمكن استخراجها لتحديد الأنماط التي يمكن أن تؤدي لوقف أنشطة الإرهابيين، وهذه ليست بالمهمة السهلة.

قضايا تقنية حديثة في الخصوصية والتحليلات:

يعتمــد أكثر مقدمي خدمات الإنترنت مثــل Google، وFacebook، وTwitter، وغيرها على تحقيــق الدخل من إجراءات المُســتخدمين. فهم يفعلون ذلك بطــرق مختلفة، غير أن كل هذه الأســاليب في النهاية تصل إلى ملف تعريف المُستَخدِم أو تفضيلاته اعتمادًا على استخدامه. ومع تزايد مُســتخدمي الإنترنت بوجهٍ عام ومُســتخدِمي الأجهزة المحمولة بوجهٍ خاص، تم تأســيس العديد من الشركات لاســتخدام التحليلات المتقدمة لتطوير ملفات تعريف المُســتخدمين اعتمادًا على اســتخدام الأجهزة، والحركة، وجهات الاتصال الخاصة بالمُســتَخدمين. وتحتوي مجلة Wall Street عــلى مجموعة ممتازة من المقــالات تحت عنــوان (WallStreetJournal.com, 2016) «What They Know». ويتــم تحديــث هذه المقالات باســتمرار لتســليط الضوء على أحدث التقنيات وقضايا الخصوصية / الأخلاقية. وقد شــملت إحدى الشركات المذكورة في هذه السلسلة Rapleaf (وهي الآن جزء من Towerdata). وتدَّعي تقنية Rapleaf أنها قادرة على توفير ملف تعريف المُسْــتَخْدِم بمجرد معرفــة عنوان البريد الإلكتروني الخاص به. ومــن الواضح أن تقنيتهم تُمكِّنهــم من جمع معلومات مهمــة. وهناك شركة أخرى تهدف إلى تحديد الأجهزة على أســاس اســتخدامها وهي شركة BlueCava، والتي اتحدت مؤخرًا مــع شركة Qualia (Qualia.com). وتقوم تقنية BlueCava الخاصة بشركة Qualia بإرفاق ملف شــخصي لتتمكن من التعرف على المُستَخدِم باعتباره فرداً واحداً أو عائلة على الرغم من أنهم قد يستخدمون العديد من التليفونات المحمولة وأجهزة الحاســب المحمولة. وتقوم كل هذه الشركات بتوظيف التقنيات مثل التجميع واســتخراج الارتباط لتطوير ملفات تعريف المُسْــتَخدِمين. ومن المؤكد أن مثل هذه التطبيقات التحليلية تثير أســئلة شــائكة حــول انتهاك خصوصية المُسْــتَخدِمين. وبالطبع، فــإن العديد من الشركات التحليلية حديثة الظهور في هذا المجال تَدَّعي احترام خصوصية المُسْتَخدِم، بيد أنه كثيرًا مــا يتم الإبلاغ عن حدوث انتهاكات. فمثلًا، كانت شركــة Rapleaf (وكما أشرنا آنفًا، أنها متحدة الآن مــع Towerdata) تجمع معلومات غير مُصَرَّح بها لمُسْــتَخْدِم من مســتخدمي Facebook

وتــم حظرها لاحقًا من Facebook. وقد ذكر عامــود في مجلة Time بقلم Joel Stein (2011) أنه بعد مرور ســاعة من إعطاء عنوان بريده الإلكــتروني إلى شركة متخصصة في مراقبة معلومات المُســتَخدمين (reputation.com)، تمكنوا بالفعل من اكتشــاف رقم الضمان الاجتماعي الخاص بــه، ويُعد هذا الرقم مفتاحًا للوصــول إلى معلومات خاصة عن المُســتَخدِم وقد يؤدي إلى سرقة الهوية. ولذا؛ فإنه ينشأ عن انتهاكات الخصوصية مخاوف من السلوك الإجرامي الذي يستند إلى معلومات المُستَخدِم، وهذا المجال يُعد مصدر قلق كبير ويحتاج إلى دراسة متأنية. وسيقوم موقع الويــب الخاص بهذا الكتــاب بتحديث التطورات الجديدة. ويُعد موقــع مجلة «What They Know» Wall Street مصدرًا ينبغي الرجوع إليه بشكل دوري. وهذه الأمثلة لا توضح فقط قوة التحليلات في التمكن من معرفة المزيد حول العملاء المُستهدفين، بل إنها أيضًا تعمل بمثابة تحذير لمحترفي التحليلات بشأن كونها حساسة تجاه الخصوصية والقضايا الأخلاقية.

وهنــاك مجال تطبيق آخر يجمع بين تأثــير تقنية المعلومات التنظيميــة، والبيانات الضخمة، وأجهزة الاستشــعار، والاهتمامات المتعلقة بالخصوصية يقوم بتحليل ســلوكيات الموظف معتمدًا عــلى البيانات التي يتم جمعها من أجهزة الاستشــعار التي يرتديها الموظفون في الشــارة. وهناك شركــة واحدة، وهــي Humanyze، ذكرت أن العديد من هذه التطبيقــات هي جزء لا يتجزأ من أجهزة الاستشعار التي يرتديها الموظفون. وتقوم هذه المستشعرات بتعقب كل تحركات الموظف. وبطبيعة الحال، فإنه ينشــأ عن ذلك قضايا خصوصية كبيرة. فهل يجب أن تكون الشركات قادرة على رصد تحركات موظفيها بهذا التدخل؟ وقد أفادت Humanyze أن تحليلاتها تقدم تقريرًا على أســاس إجمالي عملائها فقط، فلا تتم مشــاركة بيانات المُســتخدِم الفردية. وقد لاحظوا أن بعض أصحــاب العمل يرغبون في الحصول على بيانات فردية للموظفين، بيد أن عَقْدهم يمنع هذا النوع من المشاركة بشكل واضح. وعلى أي حال، فإن أجهزة الاستشعار تؤدي إلى مستوى آخر من المراقبة والتحليلات، الأمر الذي يفرض أسئلة مثيرة للاهتمام تتعلق بالخصوصية، والقوانين، والأخلاق.

من يمتلك بياناتنا الخاصة؟

مــع النمو الأخير للبيانات حســب اســتخدامنا للتقنيــة وقدرة الشركات عــلى الوصول إليها واستخراجها، فإن النقاش المتعلق بالخصوصية يؤدي أيضًا إلى السؤال الواضح حول لمن تكون ملكية بيانات المُســتَخدِم. وقد ســلط Welch (٢٠١٦) الضوء على هذه القضية في عمود Bloomberg Businessweek. وإليك مثالًا على ســيارة جديدة نســبيًا، حيث تم تجهيز الســيارة بالعديد من أجهزة الاستشعار بدءًا بأجهزة استشعار ضغط الإطارات وصولًا إلى أجهزة تتبُّع الموقع GPS والتي

يمكنها تتبع المكان الذي تذهب إليه، ومدى السرعة التي تقود بها، ومتى قمت بتغيير الطرق، وغير ذلك. وقد تعرف السيارة حتى وزن الراكب الذي انضم إلى المقعد الأمامي. وكما يشير Welch، فإن السـيارة المتصلة بالإنترنت (وهذا حال معظم السيارات الجديدة) قد تكون كابوسًا خاصًا لمالكها أو "منجم ذهب" بيانات لمن يستطيع امتلاك هذه البيانات. وهناك معركة كبيرة محتدمة بين شركات صناعـة السـيارات ومقدمي التقنيـة مثـل: Apple (Car Play)، وGoogle (Android Auto) عـلى من يمتلك هـذه البيانات ومن الذي يمكنه الوصول إليها. وهـذا الأمر أصبح أكثر أهمية لأن السـيارات أصبحت ذاتية القيادة بشكل أكبر، ومن الممكن أن يكون السائق / الراكب في السيارة عميلًا مسـتقبليًا مستهدفًا بشكل كبير لمنتجات وخدمات محددة ومعروفة بشكل جيد للمنظمة القادرة على إنشـاء ذلك الملف الشـخصي. فعلى سـبيل المثال، يجمع تطبيق Google's Waze بيانات نظام تحديد الموقع للمُسـتخدِم لما يزيد عن ٥٠ مليون مُسـتَخدِم لتتبع معلومات الحركة المرورية ومساعدة المستخدمين في العثور على أفضل مسار، ولكن عندئذٍ تُعرض الإعلانات المنبثقة على شاشات المستخدمين. ومن الجدير بالذكر أن تطبيقات Yelp وSpotify والتطبيقات الأخرى شائعة الاستخدام في السيارة لها نفس الخطط والتطبيقات.

وهناك معركة أخرى مشابهة محتدمة حول صحة المستخدمين وبيانات القياس الحيوي. وبسبب المخاوف الأمنية، فإن العديد من المسـتخدمين يتحولون إلى تسـجيل الدخول الحيوي للتحقق من الشخصية باستخدام بصمات الأصابع، وشاشات اللمس، وغير ذلك. ونظرًا لأن هذه المعلومات ينفرد بها الفرد دون غيره بشـكل كبير، فقد يصبح التنميط المستقبلي للمستخدم أكثر دقة. وبالتالي، فإن معركة امتـلاك هذه المعلومات وربطها ببيانات أخرى تم جمعها تتزايد هي الأخرى. وبالمثل، فإن المستشـفيات والأخصائيـين الطبيين والمختبرات وشركات التأمين تجمـع الكثير من المعلومات حول تاريخنا الطبي. وعلى الرغم من وجود قوانين صارمة في الولايات المتحدة (مثل: HIPAA) لحماية خصوصية المسـتخدم، إلا أن تجميع مثل هذه المعلومات يطلق العنان للتقدم الكبير في التحليلات الصحية. ومع ذلك، فإن التحدي المتعلق بالخصوصية، لايزال حقيقيًا للغاية.

ولكي تكون خبيرًا في تحليل البيانات، فإن الحد الأدنى لذلك، أن تكون مدركًا تمامًا للقضايا القانونية والأخلاقية التي ينطوي عليها جمع المعلومات التي قد تكون خاصة أو محمية. وهناك سـؤال عام لتطرحه على نفسك وهو: هل ترغب في تضمين معلوماتك الخاصة في التطبيق الذي تفكر فيه؟

الأخلاق في اتخاذ ودعم القرار:

وينقلنا السؤال الأخير إلى العديد من القضايا الأخلاقية المرتبطة بالتحليلات. وتتضمن القضايا الأخلاقية التي قد تمثل أهمية في تنفيذ التحليلات ما يلي:

- المراقبة الإلكترونية.
- الأخلاق في تصميم أنظمة دعم القرار (انظر Chae، وParadice، وCourtney، وCagle، 2005).
- سرقة البرمجيات.
- انتهاك خصوصية الأفراد.
- استخدام قواعد بيانات الملكية.
- استخدام الملكية الفكرية مثل المعرفة والخبرة.
- تعرض الموظفين لبيئات غير آمنة مرتبطة بأجهزة الحاسب.
- إمكانية وصول العاملين غير المصرح لهم إلى جهاز الحاسب.
- دقة البيانات، والمعلومات، والمعرفة.
- حماية حقوق المستخدمين.
- سهولة الوصول إلى المعلومات.
- استخدام أجهزة حاسب الشركة لأغراض ليس لها علاقة بالعمل.
- مقدار تفويض أجهزة الحاسب في اتخاذ القرار.

تشــكل القيم الشخصية عاملًا مهمًا في قضية اتخاذ القرارات الأخلاقية. وتعتبر دراسة القضايا الأخلاقيــة معقــدة نظرًا لأنها متعــددة الأبعاد. لذلك؛ فمــن المنطقي أن تتطــور الأُطر لوصف العمليــات الأخلاقية والنظــم. وقد بيّن كل من Mason، Culnan (1995) كيف توســع التقنية والابتــكار من حجم مجال الأخلاقيات وكيــف تناقش نموذجًا للتفكير الأخلاقي ينطوي على أربعة أسئلة أساسية للتركيز، وهي:

١- من هو الوكيل؟
٢- ما هو الإجراء الذي اتخذ بالفعل أو يجري التفكير فيه؟
٣- ما هي نتائج أو عواقب الفعل؟
٤- هل النتيجة عادلة أم أنها فقط لجميع أصحاب المصلحة؟

كما وصفوا التسلســل الهرمي للتفكير الأخلاقي الذي يستند فيه كل حكم أو عمل أخلاقي إلى قواعد وقوانين الأخلاق، والتي تقوم على المبادئ التي تستند بدورها على نظرية أخلاقية.

هناك قصة واحدة هي التي جعلت العديد من المســتخدمين مســتائين (على الرغم من أنها لم تكــن غير قانونيــة)، فقد كانت تجربة الـ Facebook هي إحــدى تجارب الماضي وكانت هذه

التجربة لتقديم الأخبار المختلفة للمستخدمين ومراقبة ردود أفعالهم العاطفية والتي كانت تقاس من خلال الردود، والإعجابات، وتحليل المشاعر، وغير ذلك (وعلى سبيل المثال، انظر Goel، 2014). وتعمل معظم الشركات، بما في ذلك شركات التقنية على تشغيل اختبار المُستخدِم لتحديد الميزات التي تحظى بميول الكثيرين إليها أو التي تكون غير مُحبَّبة ومن ثم تقوم الشركات بتعديل عروض منتجاتهــا. ونظرًا لأن الـ Facebook كبير للغاية، فإن أمر تشــغيل هــذه التجربة بدون الموافقة المســبقة للمستخدمين يُعد أمرًا غير أخلاقي. وفي الواقع، فقد اعترف Facebook بخطئه ومن ثَمّ أُعدت المزيد من المراجعة الرســمية من خلال لوحات المراجعة الداخلية وآليات الامتثال الأخرى للاختبار في المســتقبل. وعلى الرغم من أنهم واجهوا الكثير من الصحف السيئة في البداية، إلا أن ردهم في الوقت المناسب سمح لهم باستعادة مكانتهم سريعًا.

أسئلة مراجعة على القسم ٨-٥:

١- اذكر بعض القضايا القانونية الخاصة بالتحليلات.

٢- قم بوصف المخاوف المتعلقة بالخصوصية في التحليلات.

٣- من وجهة نظرك، من الذي ينبغي أن يمتلك البيانات المتعلقة باستخدامك للسيارة؟

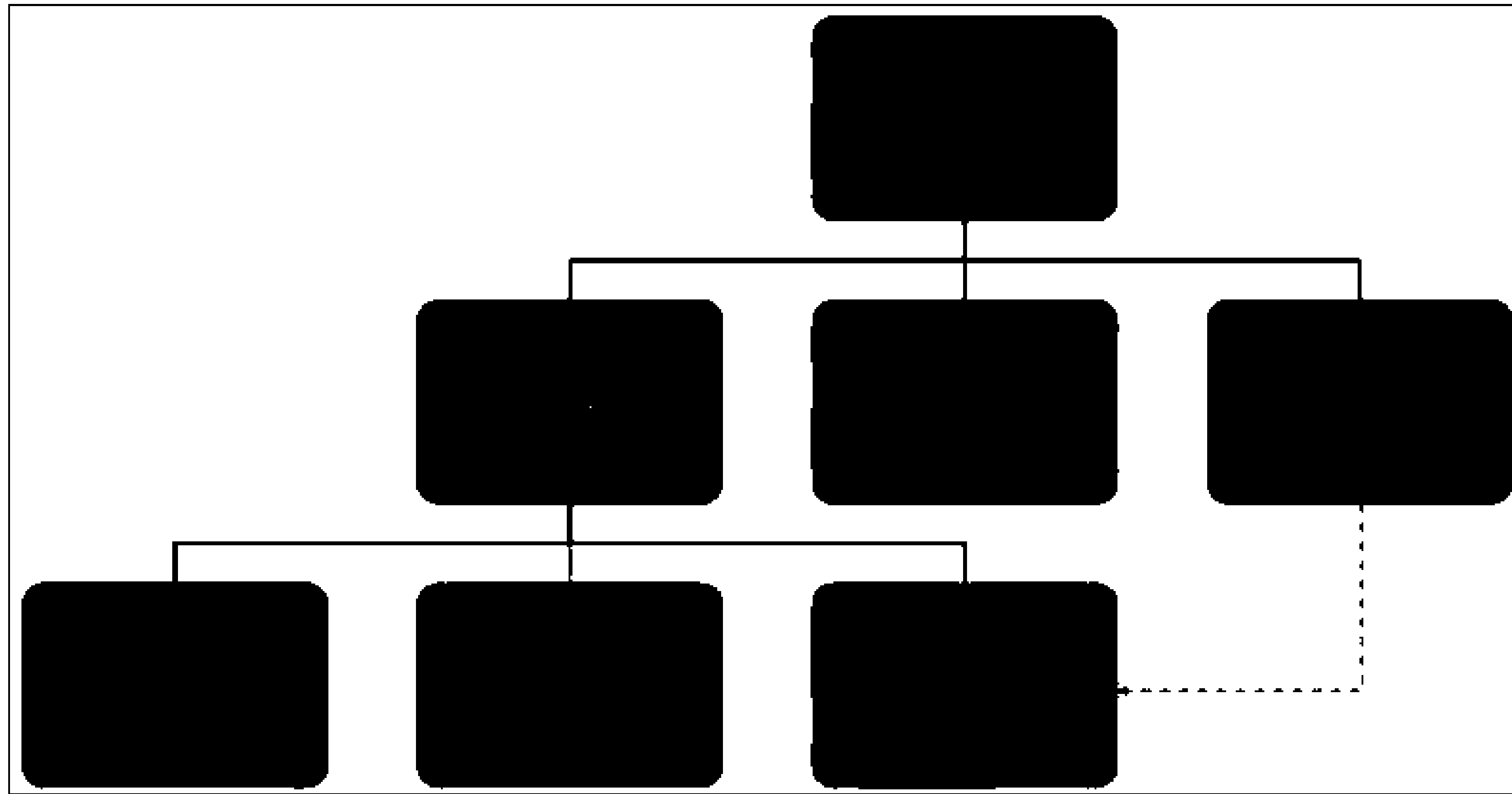
٤- اذكر القضايا الأخلاقية في التحليلات.

٨-٦ آثار التحليلات في المنظمات: نظرة عامة:

تُعد الأنظمة التحليلية عوامل مهمة في ثورة المعلومات والمعرفة، وهذا تحول ثقافي يلتقي فيه معظــم الناس الآن فقط للتوصل إلى بنــود. وعلى العكس من الثورات البطيئة التي كانت تحدث في المــاضي، مثــل الثورة الصناعية، فإن هذه الثورة تحدث بسرعــة كبيرة وتؤثر على كل جانب من جوانب حياتنا. ويلازم هذا التحول السريع مجموعة من القضايا الإدارية والاقتصادية والاجتماعية.

إن فصل أثر التحليلات عن تلك التأثيرات الخاصة بالأنظمة الأخرى المحوسبة لهي مهمة صعبة، لا سيما بسبب الاتجاه نحو الدمج، أو حتى التضمين، والتحليلات مع أنظمة المعلومات الأخرى التي تعتمد على الحاسب. ويمكن للتحليلات أن يكون لها آثار دقيقة وآثار جسيمة على حد سواء. فهذه الأنظمة يمكنها أن تؤثر على الأفراد والوظائف، بالإضافة إلى أنها يمكنها التأثير أيضًا على هياكل عمل الإدارات والوحــدات داخل المنظمــة. كما يمكن أن يكون لها أيضًا تأثيرات كبيرة على المدى الطويل على الهياكل التنظيمية الإجمالية، والصناعات بأكملها، والمجموعات، والمجتمع ككل (أي: أثر كلي).

وســوف يكون للنمو المفاجئ في التحليلات، والذكاء الاصطناعي، والحوســبة المعرفية أثر كبير على مســتقبل المنظمات. ويمكن تقســيم أثر أجهزة الحاســب والتحليلات إلى ثلاث فئات عامة: تنظيمية، وفردية، ومجتمعية. وقد كان لأجهزة الحاسب العديد من التأثيرات في كل فئة من هذه الفئات. ولا يمكننا النظر في جميع هذه الأمور في هذا القسم. لذا؛ فإننا سوف نتطرق في الفقرات التاليــة إلى الموضوعات التي نرى أنها أكثر صلة بالتحليلات. ويوضح الشــكل (٨-٧) الموضوعات العامة التي نخطط لتغطيتها.



شكل ٨-٧: أثر التحليلات على المنظمات

وحدات تنظيمية جديدة:

ويتمثل أحد التغييرات في الهيكل التنظيمي في إمكانية إنشــاء قســم للتحليلات، أو قسم ذكاء الأعمال، أو قســم علم البيانات الــذي تلعب فيه التحليلات دورًا رئيسًــا. وهذه الوحدة الخاصة يمكن دمجها مع وحدة التحليل الكمي أو اســتبدالها، أو يمكن أن تكون كيانًا جديدًا تمامًا. وبعض الــشركات الكبيرة لديها وحدات أو إدارات منفصلة لدعم القرار. فعلى ســبيل المثال، فإن العديد من البنوك الكبرى لديها مثل هذه الإدارات في أقسام الخدمات المالية الخاصة بها. كما أن العديد من الشركات لديها وحدات صغيرة من علم البيانات أو ذكاء الأعمال / مستودع البيانات. وعادةً ما تشارك هذه الأنواع من الإدارات في التدريب بالإضافة إلى أنشطة تطوير الاستشارات والتطبيقات. وهنــاك شركات أخرى فوَّضت أحد كبار مســؤولي التقنية ليكون مســؤولًا عن ذكاء الأعمال، وعن الأنظمــة الذكية، بالإضافة إلى تطبيقــات التجارة الإلكترونية. وهنــاك شركات مثل شركة Target

وشركة Walmart لديها استثمارات كبيرة في مثل هذه الوحدات، والتي تعمل باستمرار على تحليل بياناتها لتحديد كفاءة التسويق وإدارة التوريد من خلال فهم تفاعلات كلٍ من العملاء والموردين. وعلى الجانب الآخر، فإن هناك شركات عديدة تقوم بدمج تخصصات علم التحليلات / البيانات داخل المجالات الوظيفية مثل التسويق، والتمويل، والعمليات. وبشكل عام، فإن هذا المجال هو أحد المجالات التي يوجد بها فرص عمل كبيرة حاليًا. وقد حددنا في الفصل الأول عناوين الوظائف التي لها صلة بهذا الموضوع، كما سنقدم وصفًا عامًا لعالم البيانات المثالي في القسم التالي. وقد أدى نمو صناعة ذكاء الأعمال إلى تكوين وحدات جديدة داخل الشركات التي تقدم خدمات تقنية المعلومات أيضًا. فعلى سبيل المثال، قامت IBM قبل بضع سنوات، بتشكيل وحدة أعمال جديدة تركز على التحليلات. وقد تضمنت هذه المجموعة وحدات في ذكاء الأعمال، ونماذج الأمثلية، والتنقيب في البيانات، وأداء الأعمال. والأهم من ذلك، هو أن المجموعة لا تركز فقط على البرمجيات، بل إنها تركز بشكل أكبر على الخدمات / الاستشارات. وكما هو مُلاحظ في الأقسام السابقة، فإن النمو الهائل في صناعة التطبيقات قد أنشأ العديد من الفرص للشركات الجديدة التي يمكنها توظيف التحليلات وتقديم تطبيقات مبتكرة في أي مجال محدد. كما لاحظنا فيما سبق كيف تقوم الشركات الصناعية التقليدية في العصر الصناعي مثل شركة General Electric وشركة Siemens بإعادة تأهيل نفسها لتوليد عائدات كبيرة من تقديم خدمات التحليلات التي لها صلة بمنتجاتها وخدماتها. وهذا من شأنه أن يغير هيكل السلطة التنظيمية؛ نظرًا لأن هيكل القوة التنظيمية عادةً ما يُشتق من المكان الذي يتم فيه إنشاء معظم الدخل.

إعادة تصميم المُنظمة من خلال استخدام التحليلات:

يقوم مجال بحثي وعملي حديث الظهور بتوظيف تقنيات علم البيانات لدراسة الديناميكيات التنظيمية، وسلوك الأفراد، وإعادة تصميم المنظمة لتحقيق أهدافها بشكل أفضل. وتُعرف تطبيقات هذه التحليلات في الواقع باسم «تحليلات الأشخاص». فعلى سبيل المثال، تُستَخْدَم التحليلات من قِبل إدارة الموارد البشرية لتحديد المُرَشحين المثاليين من التجمع الذي يُقدم سير ذاتية للمنظمة، أو حتى من خلال تجمُّع أوسع مثل LinkedIn. وقد تم تطوير التطبيقات لتحديد أفضل المرشحين الذين لا يحتمل مغادرتهم للمنظمة. فالإبقاء على الموظف يشبه الإبقاء على العميل، ولذلك، فإن تقليل "الحركة" داخليًا لهو أمر مهم بنفس القدر لنجاح المنظمة.

ويتعلق مجال التطبيق الحديث والأكثر إثارة بفهم سلوك الموظفين من خلال مراقبة تحركاتهم داخل المنظمة واستخدام تلك المعلومات لإعادة تصميم التخطيط أو لإعادة تشكيل الفِرق لتحقيق

أداء أفضل. وهناك شركة تُسمى Humanyze (كانت معروفة سابقًا باسم Sociometric Solutions وذُكِرَت في القسم ٨-٥) لديها شارات تتضمن نظام تحديد المواقع GPS وجهاز استشعار. وعند ارتداء الموظفين لهذه الشارات، يتم تسجيل جميع تحركاتهم. وحسبما وَرَد، فقد تمكنت شركة Humanyze من مساعدة الشركات في التنبؤ بأي نوع من الموظفين من المُرَجح بقاؤه بالشركة أو مغادرتها على أساس تفاعلات هؤلاء الموظفين مع غيرهم من الموظفين الآخرين. فعلى سبيل المثال، الموظفون الذين يبقون في مقصوراتهم الخاصة تكون فرصة ترقيهم في سلم الشركات أقل بكثير من الموظفين الذين يتنقلون ويتفاعلون مع غيرهم من الموظفين الآخرين على نطاق واسع. وقد ساعد جمع وتحليل البيانات المتشابهة شركات أخرى على تحديد حجم قاعات المؤتمرات المطلوبة أو حتى تصميم المكتب لتحقيق أقصى قدر ممكن من الكفاءة. ووفقًا لما ورد بموقع Humanyze على الإنترنت، فقد أرادت إحدى الشركات فهم خصائص قادتها بشكل أفضل. ومن خلال تحليل البيانات من تلك الشارات، تمكنت الشركة من إدراك أن القادة الناجحين لديهم بالفعل شبكات أكبر يتفاعلون معها، كما أنهم يقضون وقتًا أطول في التفاعل مع الآخرين، فضلًا عن أنهم يتمتعون بالنشاط البدني. وقد استُخْدِمَت المعلومات التي جُمِعَت من جانب قادة الفِرَق لإعادة تصميم مساحة العمل والمساعدة في تحسين أداء القادة الآخرين. ومن الواضح أن هذا يؤدي إلى قضايا تتعلق بالخصوصية، غير أن مثل هذه الدراسات قد تكون قابلة للتطبيق داخل المنظمة. ويحتوي موقع Humanyze على العديد من دراسات الحالة المثيرة للاهتمام والتي تقدم أمثلة على كيفية استخدام تقنيات البيانات الضخمة لتطوير هياكل فريق أكثر كفاءة وتصميم تنظيمي.

أثر التحليلات على أنشطة المديرين، وأدائهم، ورضاهم الوظيفي:

على الرغم من أن العديد من الوظائف قد تتمتع بثراء كبير في جانب التحليلات، إلا أن هناك وظائف أخرى قد تصبح أكثر رتابة وأقل إرضاء. فعلى سبيل المثال، قبل أكثر من ٤٠ عامًا، تنبأ Argyris (1971) بأن أنظمة المعلومات التي تعتمد على الحاسب من شأنها أن تقلل من السلطة التقديرية الإدارية في عملية اتخاذ القرار وهذا يؤدي إلى عدم رضا المديرين. ومع ذلك، فقد وجد كل من Davenport و Harris (2005) في دراستهم لنظم القرار الآلي أن الموظفين الذين يستخدمون هذه الأنظمة، وخصوصًا من يتم تمكينهم بواسطة الأنظمة، كانوا أكثر رضًا عن وظائفهم. وإذا أمكن أداء العمل الروتيني والعمل البسيط باستخدام نظام تحليلي، فإنه يجب تهيئة المديرين والعاملين في مجال المعرفة لمواجهة المزيد من التحديات. ويعد أهم عمل يقوم به المديرون هو اتخاذ القرارات. وتستطيع التحليلات أن تغير الطريقة التي يتم بها اتخاذ العديد من القرارات،

وبالتالي فإنه يمكن تغيير مسـؤوليات الوظائف الخاصة بالمديرين. وعلى سـبيل المثال، فقد وجد كلٌ مـن Perez-Cascante، وPlaisent، وMaguiraga، بالإضافة إلى Bernard (2002) أن نظام دعم القرار قد استطاع تحسين أداء كل من المديرين الحاليين والمديرين الجدد فضلًا عن الموظفين الآخرين. فقد سـاعد المديرين على اكتساب المزيد من المعارف، والتجارب، والخبرة وبالتالي تعزيز جـودة صنع القرار. وقد أعلن العديد من المديرين أن أجهزة الحاسـب قد منحتهم أخيرًا الوقت للخروج من المكتب والتواجد في الساحة. كما وجدوا أيضًا أن بإمكانهم قضاء المزيد من الوقت في التخطيط للأنشطة بدلًا من إضاعة الوقت في حل المشاكل حيث إنه من الممكن تنبيههم للمشاكل المحتملة في وقت مبكر وذلك بفضل العوامل الذكية وغيرها من الأدوات التحليلية.

ويكمـن الجانب الآخر مـن التحدي الإداري في قدرة التحليلات عـلى دعم عملية صنع القرار بشـكل عام، بالإضافة إلى قدرتها على التخطيط الإسـتراتيجي والتحكم في القرارات بشـكل خاص. حيـث يمكن للتحليـلات أن تغير عملية صنع القـرار، بل يمكنها تغيير أنماط اتخـاذ القرار. فعلى سـبيل المثال، يتم الانتهاء من جمع المعلومات من أجل صنع القرار بسرعة أكبر في حالة اسـتخدام التحليلات. وتُعد نُظم معلومات المؤسسـة مفيدة للغاية في دعم الإدارة الإسـتراتيجية. وتُستخدم الآن تقنيات التنقيب في البيانات والنص والويب لتحسـين المسـح البيئـي الخارجي للمعلومات. ونتيجة لذلك؛ يسـتطيع المديرون تغيير نهجهم لحل المشاكل وتحسين قراراتهم بسرعة. ويقال إن سـتاربكس قامت مؤخرًا بتقديم مشروب قهوة جديد، واتخذت قرارها بشـأن التسعير من خلال تجربة عدة أسـعار مختلفة ورصد ردود الفعل على وسـائل التواصل الاجتماعي على مدار اليوم، وهذا يعني أن أسـاليب جمع البيانات لأحد المديرين قد تختلف الآن بشكل كبير عما كانت عليه في الماضي. وتشـير الأبحاث إلى أن معظم المديرين يميلون إلى العمل على عدد كبير من المشـاكل في وقت واحد، متنقلين من مشـكلة إلى أخرى كأنما ينتظرون المزيد من المعلومات عن مشـكلتهم الحالية (انظر Mintzberg، وLampel، وQuinn، وGhoshal، 2002). وتقوم تقنيات التحليلات بتقليل الوقت المطلوب لإتمام المهام في عملية اتخاذ القرار كما أنها تقضي على بعض فترات الانتظار التـي تعطل الإنتاج وذلك من خلال توفير المعرفة والمعلومات. وبالتالي، يعمل المديرون على عدد أقـل مـن المهام خلال كل يوم غير أنهم يكملون المزيد منها. وقد يكون تقليل وقت البدء المرتبط بالانتقال من مهمة إلى أخرى هو أهم مصدر لزيادة الإنتاجية الإدارية.

وقد يكون من بين الآثار المحتملة الأخرى للتحليلات على وظيفة المدير هو التغيير في متطلبات القيادة. والتي تعتبر الآن من الصفات القيادية الجيدة بشكل عام والتي قد تكون تغيرت بشكل كبير من خلال اسـتخدام التحليلات. فمثلًا، يُسـتبدل التواصل وجهًا لوجه بشـكل متكرر بالبريد

الإلكتروني، ومواقع الويكي، والمؤتمرات المحوسبة؛ وبالتالي، فإن الصفات القيادية التي ترجع إلى المظهر الخارجي من الممكن أن تصبح أقل أهمية.

وفيما يلي بعض الآثار المحتملة للتحليلات على وظائف المديرين:

- تتطلب عملية اتخاذ العديد من القرارات خبرة (تجربة) أقل.
- يمكن اتخاذ القرار بشكل أسرع بسبب توافر المعلومات وجعل بعض مراحل اتخاذ القرار آلية.
- يتطلب توفير الدعم لكبار المسؤولين التنفيذيين اعتمادًا أقل على الخبراء والمحللين؛ ويستطيع المديرون فعل ذلك بأنفسهم بمساعدة الأنظمة الذكية.
- تتم إعادة توزيع الطاقة بين المديرين. (كلما زادت قدرة المعلومات والتحليل التي يمتلكونها، زادت القوة لديهم).
- دعم القرارات المعقدة يجعلها أسرع في التطور، كما أنها ستكون ذات جودة أفضل.
- يتم التعجيل بالمعلومات المطلوبة لاتخاذ القرارات عالية المستوى أو حتى يتم إنشاؤها ذاتيًّا.
- قد تؤدي الآلية في القرارات الروتينية أو في مراحل عملية اتخاذ القرار (مثل، اتخاذ القرارات عبر الخطوط الأمامية واستخدام ADS) إلى إلغاء بعض المديرين.

وبشكل عام، تبين أن وظيفة المديرين المتوسطين هي الوظيفة الأكثر عرضةً لأن تكون آلية. حيث يقوم مديرو المستوى المتوسط باتخاذ قرارات روتينية إلى حدٍ ما، والتي يمكن جعلها آلية بالكامل. وبالنسبة لمديري المستويات الدنيا فإنهم لا يقضون الكثير من الوقت في اتخاذ القرار. فهم بدلًا من ذلك، يقومون بالإشراف على غير المديرين، ويديرونهم، ويحفزونهم. وقد تكون بعض قراراتهم الروتينية آلية، مثل الجدولة؛ أما القرارات الأخرى التي تنطوي على جوانب سلوكية فلا يمكن جعلها آلية. وعلى أي حال، فإننا حتى لوقمنا بجعل دورهم الأساسي آليًّا بشكل كامل، فإننا لن نتمكن من جعل وظائفهم آلية. ويوفر الويب فرصة لجعل مهام معينة لموظفي الخطوط الأمامية آلية؛ وبالتالي فإن ذلك يؤدي إلى خفض عبء العمل على المديرين المخولين بالموافقة. أما وظيفة كبار المديرين فهي أقل روتينية، وبالتالي يصعب جعلها عملية آلية.

إعادة الهيكلة الصناعية:

بدأ عدد قليل من المؤلفين بالتكهن بأثر الذكاء الاصطناعي، والتحليلات، والحوسبة الإدراكية على مستقبل الصناعة. وهناك القليل من المصادر الحديثة الممتازة التي يمكن مراجعتها وهي (2016)

Autor (2016) وRansbotham وتقرير خاص من قِبل ذي إيكونوميست (The Economist) (Standage, 2016)، وكتاب كتبه كل من Brynjolfsson وMcAfee (2016)ويُعد تقرير ذي إيكونوميست The Economist تقريرًا شاملًا إلى حد كبير، كما أنه ينظر في العديد من الأبعاد لتأثير التطورات الحالية على الصناعة والمجتمع. وتتمثل الأسباب الرئيسة في أن التقنية تُمكِّن أجهزة الحاسب الآن من القيام بالمزيد والمزيد من المهام التي قام بها البشر. وهذا بالطبع، حدث من قبل، في الثورة الصناعية. وما يجعل التغيير هذه المرة أكثر أهمية بكثير هو أن التقنية تُمكِّن الآلات من القيام بالعديد من المهام المعرفية. كما أن سرعة التغيير جذرية إلى الحد الذي يدعو إلى التنبؤ بأن الأثر المحتمل على المنظمات والمجتمع سيكون مهمًا للغاية، وفي بعض الأحيان لا يمكن التنبؤ به. وبالطبع، فإن هؤلاء المؤلفين لا يتفقون في تنبؤاتهم. وفي هذه الفقرة دعونا نركز على الآثار التنظيمية أولًا. حيث يناقش Ransbotham (2016) أن الحوسبة الإدراكية ستحول العديد من الوظائف التي يقوم بها البشر ليتم القيام بها بواسطة أجهزة الحاسب، مما يترتب عليه تقليل تكاليف المنظمات. كما أن جودة المخرجات قد تزيد أيضًا في العمل المعرفي، والتي ظهرت في العديد من الدراسات التي تقارن أداء الإنسان بأداء الآلة. فالجميع على علم تام بفوز IBM Watson في Jeopardy أو نظام Google الفائز في لعبة GO ضد الأبطال البشريين. غير أن العديد من الدراسات الأخرى في مجالات محددة مثل التعرف على الكلام وتفسير الصور الطبية قد أظهرت تفوقًا مماثلًا للنظام الآلي في حالة ما إذا كانت المهمة عالية التخصص ولكنها روتينية أو متكررة. أيضًا، فإنه نظرًا للاتجاه نحو توفير الأجهزة في جميع الأوقات وفي جميع المواقع، فإن ذلك قد يزيد من اتساع المنظمة، مما يؤدي إلى سهولة التدرج وبالتالي منافسة أكبر بين المنظمات. وهذه الآثار التنظيمية تعني أن المنظمات التي كانت بارزة بالأمس من الوارد ألا تظل في المقدمة إلى الأبد، لأن الحوسبة الإدراكية والآلية يمكنها تحدي اللاعبين الراسخة أقدامهم. وهذا هو الحال أيضًا في صناعة السيارات. وعلى الرغم من محاولة شركات السيارات التقليدية اللحاق بالركب بسرعة، إلا أن شركات كل من Google، وTesla، وشركات التقنية الأخرى تعمل على تعطيل هيكل الصناعة من خلال تحدي قادة عصر السيارات. وتقوم التحليلات بتمكين العديد من هذه التغييرات.

أثر الأتمتة على الوظائف:

كما تناقش التقارير البحثية التي سبق أن حددناها في الفقرة السابقة أيضًا تأثير التقدم في علم البيانات والذكاء الاصطناعي على الوظائف البشرية. وكما أوضحنا سابقًا، فإن العديد من مهام العاملين في مجال المعرفة يمكن الآن تنفيذها آليًا. وفي الوقت ذاته، فإن التقنية لا تؤدي دائمًا إلى تقليل عدد الأشخاص الذين يتم توظيفهم. فكما لاحظ Autor (2016)، فقد تضاعف

عدد ماكينات الصرف الآلي (ATM) أربعة أضعاف من حوالي ١٠٠,٠٠٠ إلى ٤٠٠,٠٠٠ في الفترة بين عامـي ١٩٩٥ و٢٠١٠، إلا أن عدد موظفي البنـوك قد زاد في نفس الوقت من حوالي نصف مليون موظف في عام ١٩٨٠ إلى حوالي ٥٥٠,٠٠٠ موظف في عام ٢٠١٠. فما حدث هو أن الجزء الروتيني مـن العمل أصبح يتم عن طريق أجهزة الصرف الآلي (ATM)، أما موظفو البنك فقد أصبحوا على اتصـال أفضـل بالعملاء كما أصبحوا يركـزون الآن على خدمات البيـع. وبالطبع، فإن بعض هذه الخدمات يمكن أيضًا تقديمها من خلال نماذج التحليلات التي تحدد العملاء المناسبين للحصول على فرصة محددة. وقد قدم Ransbotham (2016) مثالًا آخر. وعادةً ما تعتبر الاستشـارات المالية مهمة تعتمد على المعرفة. ونظرًا لأن تقنية علم البيانات توفر دعمًا مخصصًا لسيناريو محدد، فإن تكاليف هذه الخدمات سوف تنخفض. وهذا سيؤدي إلى تزايد الأشخاص الذين يختارون المطالبة بمثل هذه الخدمات، مما سيؤدي في النهاية إلى الحاجة لوجود المزيد من البشر للعمل المتقدم.

ويدعي بعض هؤلاء المؤلفين بأن الأتمتة التي ترجع إلى الحوسـبة المعرفية والذكاء الاصطناعي سـوف تسرع ما يسـمى بـ «الاستقطاب» لسـوق العمل في المسـتقبل. وهذا يتبعه نمو كبير في الوظائف في المستويات العليا والدنيا من سوق العمل، غير أن الخسائر تكون في الوسط. وتستمر الوظائف التي تتطلب مهارات منخفضة كخدمات الحراسـة، والرعاية الشخصية، وإعداد الطعام، وما إلى ذلك، في النمو. وبالمثل، فإن الوظائف التي تتطلب مستويات مهارة عالية جدًا، كالتصميم الإداري، وتصميم الرسـومات، والعمل الحاسـوبي تتزايد هي الأخرى، في حـين أن الوظائف التي تتطلب معرفة متخصصة "المهارات المتوسطة" والتي تم تطبيقها مرارًا وتكرارًا مع بعض التكيف، هي الوظائف الأكثر عرضةً للاختفاء. ففي بعض الأحيان تتداخل التقنية في نفسـها. فعلى سبيل المثال، فإن تحليلات IBM Watson تتضمن الآن إمكانات أولية للبدء في طرح الأسئلة التي يمكن أن يطلبها أحد متخصصي التحليلات من مخزن البيانات، ومن الواضح، أنها تُقدم إجابات. وبالمثل، فإن التحليلات الأخرى - كعروض الخدمات - قد تؤدي إلى تقليل عدد الأشخاص الذين يحتاجون إلى الكفاءة عند استخدام برامج التحليلات.

ويشـير تقرير ذي إيكونوميسـت إلى أنه حتى لـو لم يحل الذكاء الاصطناعـي محل العمال بشـكل مباشر، فإن الأمر سـوف يتطلب منهم بالتأكيد اكتسـاب مهارات جديدة ليظلوا قادرين على المنافسـة. كما أن تعطل السـوق دائمًا ما يكون غير مريح. وسـوف توفر السـنوات القليلة المقبلة فرصًا ممتازة لمتخصصي التحليلات لتشكيل المستقبل. وفي القسم التالي سوف نحدد بعض السمات العامة لمتخصص علم البيانات.

الآثار غير المقصودة للتحليلات:

إلى جانب القضايا التي ناقشناها بالفعل فيما يتعلق بالخصوصية، والأخلاقيات، والأمان، والآثار الشخصية / التنظيمية للتحليلات، يجب أن يكون المديرون والمختصون بعلم البيانات على دراية بالآثار الاجتماعية طويلة الأجل للنماذج. وقد قدم كتاب صدر مؤخرًا لـ Cathy O'Neil (2016) هذه الحجج بشكل جيد للغاية مما لاقى انتشارًا في الأخبار. وقد حصلت O'Neil على الدكتوراه في الرياضيات من جامعة هارفارد Harvard وعملت في مجال التمويل وصناعة علم البيانات. وقد دفعتها تجاربها وملاحظاتها إلى كتابة كتاب مشهور بعنوان أسلحة تدمير الرياضيات Weapons of Math Destruction: كيف تزيد البيانات الضخمة من عدم المساواة وتهدد الديموقراطية. ونحن ندعوك لقراءة الكتاب، أو على الأقل الاطلاع على موقع مدونتها عبر /https://mathbabe.org. وإلى جانب الترويج لكتابها، فإن موقع المدونة يسلط الضوء على القضايا الاجتماعية التي تتعلق بالتحليلات. كما يمكنك الاطلاع على ملخص / استعراض جيد للكتاب عبر هذا الموقع: //:http .knowledge.wharton.upenn.edu/article/rogue-algorithms-dark-side-Big-data/

وتعتقد O'Neil في كتابها، بأن النماذج يجب أن تستوفي ثلاثة شروط. فالعديد من النماذج الرياضية ليست شفافة. فإذا كان النموذج غير مفهوم، فإن تطبيقه يمكن أن يؤدي إلى عواقب غير مقصودة. ثانيًا، يجب أن يتسم النموذج بأهداف واضحة قابلة للقياس الكمي، مثل التطبيق المشهور للتحليلات في كتاب وفيلم كرة المال Moneyball، وقد كان الهدف من هذا النموذج هو زيادة عدد مرات الفوز. وقد كانت مقاييس المدخلات المقترحة من العمر يمكن فهمها كذلك. بدلًا من استخدام قاعدة التشغيل المعلنة والأكثر شيوعًا في قياس (RBI)، والتي اقترحها المُحَلل واستخدمت في النسب الأساسية والتدابير الأخرى (التي تم حسابها بسهولة وفهمها بسهولة من قِبل أي شخص لديه مهارات الرياضيات الأساسية). ومن ناحية أخرى، فإن النماذج التي بُنيت لتقييم مخاطر الأوراق المالية المدعومة بالرهن العقاري حيث لا يفهم أحد تمامًا الافتراضات الأساسية سوى المتداولين الماليين؛ حيث كانوا يتداولون الأوراق المالية المضمونة على نطاق واسع وذلك لأنهم قادوا الأزمة المالية في عام ٢٠٠٨. أما الشرط الثالث فهو أن النماذج يجب أن تكون لها آلية تصحيح ذاتي وعملية في المكان بحيث يتم تدقيق النماذج بانتظام، كما يجري النظر باستمرار في المدخلات والمخرجات الجديدة. وهذه المسألة الثالثة تُعد بالغة الأهمية في تطبيق النماذج في الأماكن الاجتماعية. ومن ناحية أخرى فإن النماذج تديم الفرضيات الخاطئة الكامنة في مرحلة النمذجة الأولية. وتناقش O'Neil العديد من المواقف في هذه الحالة. فمثلًا، تُستَخدم في الولايات المتحدة لتحديد المدرسين ذوي الأداء الضعيف ومكافأة المعلمين الأفضل. وقد استخدمت هذه

النــماذج درجات اختبار تلاميذهم. وتروي O'Neil العديد من الأمثلة التي اســتخدمت النماذج لتسريح المعلمين «ضعيفي الأداء» على الرغم من أن هؤلاء المعلمين كانوا محبوبين من قِبل الطلاب وأولياء الأمور. وتستشــهد O'Neil أيضًا بمثال آخر تتزايــد أهميته في العديد من المنظمات، وهو أداء العمال؛ حيث يتم اســتخدام النماذج لتحســين جدولة العمال في العديد من المنظمات. وفي العديــد من الحالات، يتم تطويــر هذه الجداول لتتوافق مع تغيرات الطلب الموســمية واليومية، إلا أن O'Neil تأســف لحقيقــة أن هذه النــماذج لا تأخذ في الاعتبار الآثار الضــارة لهذا التفاوت في الجــداول على أُسر هــؤلاء العمال ذوي الدخل المنخفض. وهنــاك أمثلة أخرى من هذا القبيل تشــمل نماذج تقييم درجة الائتمان والتي تســتند على ملامح تاريخية وبالتالي فإنها قد تؤثر سلبًا على الأقليات. وبدون وجود آليات لتدقيق هذه النماذج وآثارها غير المقصودة، فإننا قد نضر أكثر مما ننفع على المدى الطويل. ولذلك؛ فإن متخصص علم البيانات يحتاج إلى إدراك هذه المخاوف.

أسئلة مراجعة على القسم ٨-٦:

١- اذكر آثار التحليلات على اتخاذ القرار.

٢- اذكر آثار التحليلات على المهام الإدارية الأخرى.

٣- قم بوصف الوحدات التنظيمية الجديدة التي أُنشئت بسبب التحليلات.

٤- حدد أمثلة أخرى لتطبيقات التحليلات لإعادة تصميم مساحة العمل أو سلوك الفريق.

٥- كيف تؤثر الحوسبة الإدراكية على هيكل الصناعة؟

٦- ما هي الوظائف التي من المرجح أن تتغير نتيجة التحول للآلية؟

٧- ادرس تقريــر ذي إيكونوميســت (Standage, 2016) المذكور في هذا القســم. ما هي الآثار الأخرى للآلية؟ وهل وجدتها مثيرة للاهتمام؟

٨-٧ عالم البيانات كمهنة:

عــالم البيانات هو دور أو وظيفة مرتبطــة في كثير من الأحيان بالبيانات الضخمة. وقد أصبح هــذا الدور من أكثر الأدوار المطلوبة في الســوق. وفي مقــال نُشر في عدد أكتوبر ٢٠١٢ من مجلة Harvard Business Review، أطلــق العالمــان Thomas H. Davenport، وD. J. Patil على وظيفة عالم البيانات بأنها "الوظيفة الأكثر جاذبية في القرن ٢١". وقد حددوا في تلك المقالة علماء البيانات الأساسيين، والأكثر مهارة عالمية، إضافة إلى من لديهم القدرة على كتابة التعليمات البرمجية

(بأحدث لغات ومنصات البيانات الضخمة). وعلى الرغم من أن هذه المهارات المذكورة قد تصبح في المستقبل القريب أقل دقة خاصة عندما يحمل العديد من الأشخاص لقب «عالم البيانات» على بطاقات العمل الخاصة بهم، غير أنها تبدو في وقتنا الراهن بأنها أكثر المهارات الأساســية المطلوبة من علماء البيانات. وستكون أكثر المهارات استمرارية هي حاجة علماء البيانات إلى التواصل بلغة يفهمهــا جميع أصحاب المصلحة، بالإضافة إلى حاجتهم لإظهار المهارات الخاصة التي تتضمن سرد القصص مع البيانات، سواء أكان ذلك شفهيًّا و / أو بصريًّا، (Davenport, Patil, 2012).

يســتخدم علــماء البيانات مزيجًا من مهاراتهــم الخاصة بالأعمال بالإضافــة إلى مهاراتهم الفنية لدراسة البيانات الضخمة، بحثًا عن طرق لتحسين ممارسات تحليلات الأعمال الحالية (من وصفي إلى تنبُّؤي وتوجيهي) ومن ثم تحســين القرارات من أجل فرص الأعمال الجديدة. ومن أكبر الاختلافات بين عالم البيانات ومُستَخدِم ذكاء الأعمال مثل محلل الأعمال، هو أن عالم البيانات يدرس ويبحث عن قدرات جديدة، في حين أن مُستَخدِم ذكاء الأعمال يقوم بتحليل حالات الأعمال والعمليات الحالية.

كما يُعد الفضول الشديد هو أحد الصفات الغالبة والمتوقعة من علماء البيانات، بمعنى وجود رغبــة عند علماء البيانات للتعمق في المشــكلة وعدم تناولها بســطحية، بالإضافة إلى العثور على الأسئلة في صميمها، واستخلاصها إلى مجموعة شديدة الوضوح من الفرضيات التي يمكن اختبارها. وهذا يســتلزم في كثير من الأحيان التفكير الجماعــي الذي يميز العلماء الأكثر إبداعًا في أي مجال. فعلى ســبيل المثال، نحن نعرف عالم بيانات يدرس مشــكلة الاحتيال والذي أدرك أنها تتشابه مع نوع من المشــاكل الخاصة بتسلســل الحمض النووي، (Davenport, Patil, 2012). وقد تمكن هذا العالم هو وفريقه من صياغة حل يقلل من خســائر الاحتيال بشــكل كبير، وذلك من خلال الجمع بين تلك العوالم المتفاوتة.

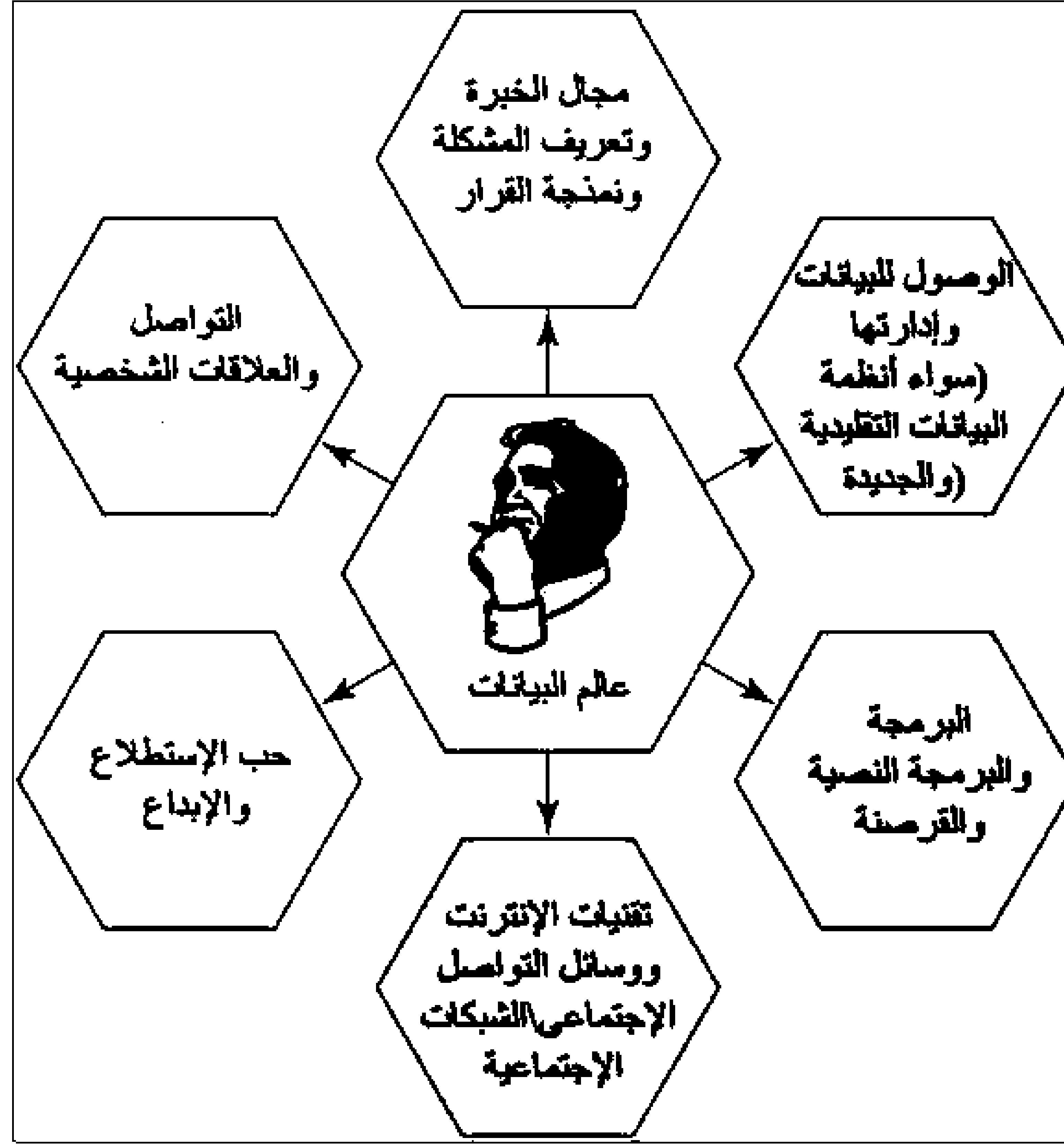
من أين يأتي علماء البيانات؟

على الرغم من وجود بعض الخلاف حول استخدام العلم في الاسم، غير أن هذا الأمر أصبح أقل إثارة للجدل. ويســتخدم علماء حقيقيون أدوات صنعها علماء آخرون، كوسيلة لتوسيع المعرفة، أو يقومون بصنعها إذا لم تكن موجودة. وهذا هو بالضبط ما يُتوقَّع القيام به من قِبل علماء البيانات. فعلى ســبيل المثال، ينبغي على علماء الفيزياء التجريبية تصميم المعدات، وجمع البيانات، وإجراء تجارب متعددة لاكتشاف المعرفة وتوصيل نتائجها. وعلى الرغم من أن علماء البيانات قد لا يرتدون معاطف بيضاء فضلًا عن أنهم قد لا يعيشــون في بيئة معمليــة معقمة، إلا أن ما يفعلونه بالضبط

يتمثل في: استخدام الأدوات والأساليب الإبداعية لتحويل البيانات إلى معلومات قابلة للتطبيق لكي يستخدمها الآخرون لاتخاذ قرارات أفضل.

الجدير بالذكر أنه لا يوجد إجماع حول الخلفية التعليمية التي ينبغي أن تكون لدى عالم البيانات. غير أنه عادةً ما يكون المرشح لذلك من أمثال: حاملي ماجستير (أو دكتوراه) في علوم الحاسب الآلي، أو نظم المعلومات الإدارية (MIS)، أو الهندسة الصناعية، أو درجات الدراسات التحليلية المشهورة الحديثة. وعلى الرغم من أن كل ذلك قد يكون ضروريًا إلا أنه ليس كافيًا لكى يطلق على أحد الأشخاص أنه عالم بيانات. فمن أهم الخصائص التي ينبغي توافرها في عالم البيانات هي الخبرة الفنية فضلًا عن الخبرة بمجالات تطبيقات الأعمال. وبهذا المعنى، فإن هذا الدور يشبه الأدوار المهنية للمهندس أو الأدوار المهنية لإدارة المشاريع، حيث يتم تقييم الخبرة بمقدار (إن لم يكن أكثر من) المهارات التقنية والخلفية التعليمية. ولن تكون مفاجأة كبيرة أن نرى في غضون السنوات القليلة القادمة شهادة مصممة لعلماء البيانات بشكل خاص (ربما يُطلق عليها اسم "محترف علم البيانات" أو "DSP" للاختصار).

ونظرًا لأن هذا الدور هو مهنة لمجال لا يزال قيد التعريف، فإن العديد من ممارسات هذه المهنة لاتزال تجريبية كما أنها بعيدة عن أن تكون موحدة. وبالتالي، فإن الشركات تكون شديدة الحساسية فيما يتعلق ببُعْد الخبرة لعالم البيانات. وبنضج هذه المهنة، والممارسات الموحدة، فإن التجربة ستكون أخف من المشكلة عند تعريف عالم البيانات. وفي الوقت الحاضر، فإن الشركات تبحث عن الأشخاص ذوي الخبرة الواسعة في العمل مع البيانات المعقدة الذين حظوا بالتوظيف الجيد من بين أولئك الذين لديهم خلفيات تعليمية وعملية في العلوم الطبيعية أو الاجتماعية. وقد كان بعض أفضل وألمع علماء البيانات حاصلين على الدكتوراه في مجالات مقصورة على فئة معينة مثل علم البيئة وبيولوجيا النظم (Davenport, Patil, 2012). وعلى الرغم من عدم وجود إجماع حول المكان الذي يأتي منه علماء البيانات، إلا أن هناك فهمًا مشتركًا للمهارات والصفات المتوقع توافرها في عالم البيانات. ويوضح الشكل (٨-٨) رسمًا بيانيًّا عالي المستوى لهذه المهارات.



شكل ٨-٨: المهارات التي تحدد عالم البيانات

فمــن المُتوقع أن يكون لــدى عالم البيانات مهارات شــخصية مثل الإبــداع، والفضول، وتبادل المعلومات / العلاقات بين الأشخاص، والخبرات الخاصة بالمجال، وتعريف المشكلة، والإدارة (تظهر في أشــكال سداسية خضراء الخلفية في أعلى الجانب الأيسر من الشكل) فضلًا عن مهارات فنية سليمة مثل معالجة البيانات، والبرمجة / القرصنة / البرمجة النصية، والإنترنت ووسائل التواصل الاجتماعي / تقنيات الشبكات (تظهر في أشكال سداسية ذات خلفية رمادية أسفل يمين الشكل). وتدور رؤية فنية (٨-١) حول الإعلان عن وظيفة نموذجية لعالم البيانات.

رؤية فنية ٨-١

وظيفة نموذجية لعلماء البيانات

تبحث [بعض الشركات] عن عالم بيانات لينضم إلى فريق تحليلات البيانات الضخمة لدينا. ومــن المتوقع أن يكون كل فرد في هذا الدور مســتعدًا للعمل كمهندس برامج، وباحث كمي. وســوف يكون للمرشــح المثالي اهتمام كبير في دراســة شــبكة اجتماعية عبر الإنترنت وشغف لتحديد الأسئلة التي تساعدنا في تشكيل أفضل المنتجات بالإضافة إلى الإجابة عن هذه الأسئلة.

المسؤوليات:

- العمل عن كثب مع فريق هندســة المنتجات لتحديد الأســئلة المهمة عن المنتج والإجابة عليها.

- الإجابة على الأســئلة المتعلقة بالمنتجات باستخدام الأســاليب الإحصائية المناسبة على البيانات المتاحة.

- نقل النتائج لمديري ومهندسي الإنتاج.

- إدارة مجموعة البيانات الجديدة وتحسين مصادر البيانات الموجودة.

- تحليل وتفسير نتائج تجارب المنتج.

- تطوير أفضل الممارسات للأجهزة والتجريب ونقل ذلك إلى فِرق هندسة المنتجات.

المتطلبات:

- الحصول على ماجستير أو دكتوراه في أحد المجالات الفنية ذات الصلة، أو أن يكون لديه أكثر من ٤ سنوات من الخبرة في دور له صلة بهذا المجال.

- خبرة واسعة في حل المشكلات التحليلية باستخدام الأساليب الكمية.

- التعــاون في المعالجة البارعة وتحليــل البيانات المعقدة ذات الحجم الكبير والبعد العالي من مصادر مختلفة.

- وجود شغف قوي للبحث التجريبي والإجابة عن الأسئلة الصعبة باستخدام البيانات.

- منهج تحليلي مرن يسمح للنتائج بمستويات متفاوتة من الدقة.

- القدرة على توصيل التحليل الكمي المعقد بطريقة واضحة، ودقيقة، وقابلة للتطبيق.

- الطلاقة في استخدام لغة برمجة نصية واحدة على الأقل مثل Python أو PHP.

- الإلمام بقواعد البيانات العلائقية وSQL.

- الخبرة المعرفية بأداة تحليل مثل R، أو Matlab، أو SAS.

خــبرة في العمل مع مجموعات كبيرة من البيانات، بالإضافة إلى الخبرة في العمل مع أدوات الحوسبة المقسمة زيادةً إلى (MapReduce، Hadoop، Hive، إلخ).

ونادرًا ما يكون هناك أشخاص يتمتعون بهذه المجموعة من المهارات، مما يفسر سبب النقص في عـدد علماء البيانات. وبسـبب زيادة الطلب على هذه الفئة من الأفراد الأقل نسـبيًا، تكون رواتب البداية لعلماء البيانات أعلى بكثير من ستة أرقام (بالدولار الأمريكي). أما الأشخاص الذين لديهم خبرة واسـعة، بالإضافة إلى أنهم يملكون خبرة في مجـال معين، فإن رواتبهم أعلى من ذلك بكثير. وفي معظم المنظمات، سـيكون من الضروري بناء فريق من الأشـخاص الذين لديهم هذه المهـارات بشـكل جماعي بدلًا من البحث عن أفـراد يتمتعون بهذه القـدرات، وفيما يلي بعض التصريحات الحديثة حول علماء البيانات:

- يقـوم علماء البيانات بتحويل البيانات الضخمة إلى قيمة ضخمة، حيث يقدمون المنتجات التي تُرضي المستخدمين كما يُقدمون الأفكار التي تُخبر بقرارات الأعمال.
- إن عالم البيانات ليس بارعًا في التعامل مع البيانات فحسـب، بل إنه يُقدر أيضًا البيانات نفسها كأصل لا يُقدر بثمن.
- بحلول عام ٢٠٢٠ سـيكون هناك ٤,٥ مليون وظيفة جديدة لعلم البيانات، والتي لن يتم شَـغْل سوى ثلثها فقط نظرًا لعدم توافر الأشخاص لشَغْل هذه الوظائف.
- علماء البيانات اليوم هم أساطير الأسواق المالية في الثمانينيات.

إن اسـتخدام علماء البيانـات لا يقتصر على شركات الإنترنت ذات التقنيـة العالية. فالعديد مـن الشركات التي لا تمتلك الكثير من التواجد على الإنترنت هي أيضًا مهتمة بمتخصصي تحليلات البيانـات الضخمة المؤهلـين تأهيلًا عاليًا. وفي الواقـع، فإن ذلك هو المجال الذي يشـهد الكثير مـن النمـو. وبالطبع، وكما لوحـظ في الكتاب عدة مـرات، فإن فرص تنظيم المشـاريع لا تزال موجودة لتطوير التطبيق التالي «القاتل» والذي من شـأنه أن يُحدِث فرقًا كبيرًا لإحدى المنظمات والمستفيدين منها وبالتالي تجعلك غنيًا ومشهورًا.

أسئلة مراجعة على القسم ٨-٧:

١- ما هو عالم البيانات؟ وما سبب زيادة الطلب عليه؟

٢- ما هي الخصائص المشتركة لعلماء البيانات؟ وأي من هذه الخصائص يُعد أكثرها أهمية؟

٣- من أين يأتي علماء البيانات؟ وما هي الخلفيات التعليمية التي لديهم؟

٤- في رأيك، ما هو الطريق لكي تصبح عالمًا عظيمًا في البيانات؟

تلخيص لأهم نقاط الفصل:

- تستطيع البيانات الجغرافية أن تعزز تطبيقات التحليلات عن طريق دمج معلومات المكان.
- يمكن التنقيب في معلومات موقع المستخدمين الحالي لتطوير الحملات الترويجية التي تستهدف مستخدم معين في الوقت المناسب.
- يمكــن اســتخدام معلومات المــكان من أجهزة الجوال لإنشــاء ملفات تعريف لســلوك وحركة المستخدم. تُمكِّن مثل هذه المعلومات الخاصة بمكان المستخدمين من إيجاد أناس آخرين ذوي اهتمامات ومعلنين متشابهين لتكييف عروضهم الترويجية.
- تســتطيع أيضًا التحليلات المبنية على المكان أن تفيد المســتهلكين بشــكل مباشر وليس الشركات فقط. ويتم تطوير تطبيقات الجوال لتمكين مثل هذه التطبيقات للتحليلات الابتكارية.
- يعتبر إنترنت الأشــياء (IoT) المنحنى الأســاسي القادم لنمو علم التحليلات / البيانات. ويطلق عليه البعض أيضًا الإنترنت الصناعي.
- تنمو تطبيقات (IoT) في كل صناعة بما في ذلك التطبيقات الطبية والتصنيع والطاقة والملاحة الجوية.
- تعتــبر إحدى الفئات الشــائعة لتطبيقــات (IoT) هي الصيانة التنبؤيــة لتخفيض وقت توقف المعدات غير الضروري وإجراء الصيانة / الإصلاحات المطلوبة قبل استلزامها.
- تقدم الحوســبة الســحابية إمكانية اســتخدام البرمجيات والأجهزة والمنصات والبنية التحتية، كلٌّ بحسب اشتراكه. كما تمكن الحوسبة السحابية المستخدم من استثمار أكثر قابلية للتطوير.
- تقدم الخدمات التحليلية القائمة على الحوســبة الســحابية إلى المنظمات أحدث التقنيات دون استثمار معنوي مسبق.
- تســتطيع التحليلات التأثير على المنظمات بطرق كثيرة، كأنظمة مستقلة أو متكاملة فيما بينها، أو مع أنظمة معلومات أخرى قائمة على الحوسبة.
- يتفاوت تأثير التحليلات على الأفراد، فقد يكون إيجابيًا أو محايدًا أو سلبيًا.
- قــد يصاحب تقديم أنظمة ذكية ظهــور قضايا قانونية جديّة، وتعتبر المســؤولية والخصوصية مجالَي المشاكل المهيمنة.
- يمكــن توقع الكثير من الآثار الاجتماعيــة الإيجابية من التحليلات، والتي تتراوح من توفير فرص للمعاقين إلى قيادة الحرب ضد الإرهاب. ومن المحتمل تحســن جودة الحياة في كل من العمل والمنزل، كنتيجة للتحليلات. وبالطبع، يكون هناك أيضًا قضايا تحتاج للاهتمام.

- إن نمو التحليلات سوف يؤدي إلى تغيرات جوهرية في هيكل الصناعة ومستقبل التوظيف.
- تكمن المعركة الكبرى حول من يمتلك جميع بيانات المســتخدم التي يتم إنشاؤها من استعمال الهواتف الذكية والسيارات وما إلى ذلك.

مصطلحات أساسية:

الحوسبة السحابية	نظم المعلومات الجغرافية (GIS)	إنترنت الأشياء (IoT)
التعريف بترددات الراديو (RFID)	الخصوصية	

أسئلة للمناقشة:

١- ما هي الفوائد المحتملة من استخدام البيانات الجغرافية المكانية في التحليلات؟ مع ضرب أمثلة.

٢- ما هي أنواع التطبيقات الجديدة التي يمكن أن تظهر من معرفة المواقع الفورية للمستخدمين؟ وماذا لو كنت تعرف أيضًا ماذا يحملون في سلة تسوقهم مثلًا؟

٣- كيف يمكن أن يستفيد المستهلكون من استخدام التحليلات، لا سيما القائمة منها على معلومات المكان؟

٤- ناقش العبارة التالية: «يعتبر إنشــاء ملفات التعريف القائم على تتبع المكان قويًا ولكنه أيضًا يضع تهديدات للخصوصية».

٥- هل تعتبر الحوسبة السحابية «مجرد منتج قديم أعيد تغليفه بعبوة جديدة»؟ وكيف تتشابه مع المبادرات الأخرى؟ وكيف تختلف؟

٦- ناقش العلاقة بين أجهزة الجوال والشبكات الاجتماعية.

٧- يدّعي البعض أن التحليلات بشكل عام تفرغ الأنشطة الإدارية من الإنسانية، ويعترض آخرون على هذا الادعاء. ناقش براهين كل من وجهتَي النظر هاتين.

٨- إن تشــخيص العدوى ووصف الأدوية هما نقاط ضعف العديد من الأطباء الممارسين (وفقًا لـ E. H. Shortliffe، أحد مطوري MYCIN). ولذا؛ يبدو أنه ســيكون من الأفضل للمجتمع أن يتم استخدام أنظمة تشخيص مبنية على التحليلات من قِبل المزيد من الأطباء. أجب عن الأسئلة التالية:

١- في رأيك لماذا يتم استخدام مثل هذه الأنظمة بالحد الأدنى من قِبَل الأطباء؟

٢- افترض أنك مدير مستشـفى ويتبع لك الأطباء الذين يتقاضون رواتبهم منها. ماذا ستفعل لإقناعهم باستخدام النظام؟

٣- إذا كانـت الفوائد المحتملة للمجتمع عظيمة جدًا، فهل يسـتطيع هذا المجتمع أن يفعل شيئًا يزيد من استخدام الأطباء لمثل هذه النظم التحليلية؟

٩- اذكر بعض مخاوف الخصوصية الرئيسة المتعلقة باستخدام التحليلات على بيانات الجوال؟

١٠- اذكر حالات جديدة للتعديات على خصوصية المستخدم وتأثيراتها على علم البيانات كمهنة.

تمارين: شبكة جامعة تيراداتا (TUN) وغيرها من التمارين اليدوية:

- اذهب إلى teradatauniversitynetwork.com وابحث عن دراسـات الحالة. قم بمشـاهدة فيديو BSI عن حالة الركاب الذين تم انقطاع الاتصال بهم والذين قد تكون رأيتهم أثناء قراءة الفصل الأول. ما هي التطبيقات الجديدة التي يمكنك تخيلها بمستوى البيانات التفصيلية التي يمكن لشركات الطيران التقاطها اليوم؟

- عـلى teradatauniversitynetwork.com، اذهـب إلى مكتبـة الحلقـات الصوتيـة. قم بإيجاد حلقات صوتية عن ذكاء الأعمال واسـعة الانتشـار والمقدمة من Hugh Watson. قم بتلخيص النقاط المعروضة من قِبَل المتحدث.

- اذهـب إلى teradatauniversitynetwork.com وابحـث عن مقاطع فيديو BSI. اسـتعرض هذه المقاطع وأجب عن أسئلة الحالة المرتبطة بها.

- توفر المجموعات المبنية على تتبع المكان احتمال الخدمات الشخصية ولكن يتحدى الخصوصية. قم بتقسيم الصف إلى فريقين للتناظر أحدهما «مع» والآخر «ضد» هذه التطبيقات.

- حدد القضايا الأخلاقية المرتبطة باتخاذ القرارات الإدارية. ابحث على الإنترنت، وانضم إلى مجموعات / المدونات المناقشة، واقرأ المقالات من الإنترنت، ثم قم بإعداد تقرير عن النتائج الخاصة بك.

- ابحث على الإنترنت لإيجاد أمثلة لكيفية قيام نظم التحليلات بتسـهيل أنشـطة مثل التفويض، والتخصيص الشامل، والعمل الجماعي.

- ابحث في مركز موارد التقنية الخاص بجمعية Bar الأمريكية (abanet.org/tech/ltrc/techethics.html) وموقع nolo.com. ما هي الاهتمامات القانونية والمجتمعية الرئيسة والمتقدمة التي تمت معالجتها هناك؟ وكيف يتم التعامل معها؟

- قم باستكشاف عدة مواقع مرتبطة بالرعاية الصحية (مثل WebMD.com وwho.int). أوجد القضايا المتعلقة بالتحليلات والخصوصية، ثم اكتب تقريرًا عن الكيفية التي تعمل بها هذه المواقع على تحسين الرعاية الصحية.
- قم بزيارة مدونة IoT Ecosystem الخاصة بـ Matt Turck على http://mattturck.com/2016/03/28/2016-iot-landscape/. حدد على الأقل شركتين فيما لا يقل عن قطاعين رأسيين وقطاعين أفقيين. اذهب إلى مواقع الويب الخاصة بهاتين الشركتين وقم بإعداد تقرير عن عروضهما ونجاحات العملاء.
- ادخل على YouTube.com وابحث عن مقاطع فيديو عن الحوسبة السحابية، وقم بمشاهدة اثنين على الأقل من هذه المقاطع. لخِّص النتائج الخاصّة بك.
- ادخل على Pandora.com واكتشف كيف تستطيع إنشاء ومشاركة موسيقى مع الأصدقاء. استكشف كيف يحلل الموقع تفضيلات المستخدم.
- ادخل على Humanyze.com واستعرض دراسات حالة متنوعة وقم بتلخيص أحد التطبيقات المثيرة من أجهزة الاستشعار في فهم التبادلات الاجتماعية في المنظمات.
- يعد الهدف من هذا التمرين هو تعريفك بإمكانيات الهواتف الذكية لتحديد النشاط البشري. وتكون مجموعة البيانات متاحة على: archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones.
- إنها تحتوي على قراءات التسارع وضابط الاتجاهات (البوصلة الدوارة) لثلاثين فردًا يضعون الهاتف الذكي في حزام يلتف حول وسطهم. تكون البيانات متاحة في صورتها الخام وتتضمن بعض جهود تجهيز البيانات، ويكون هدفك هو تحديد وتصنيف هذه القراءات إلى أنشطة، كالمشي والجري وتسلق الجبال، وما إلى ذلك. ويتوفر المزيد من المعلومات عن مجموعة البيانات في صفحة التنزيل. ويمكنك استخدام التجميع من أجل الاستكشاف المبدئي واكتساب فهم البيانات. كما يمكنك استخدام أدوات مثل R لتجهيز وتحليل هذه البيانات.

المراجع:

- Alteryx.com. (n.d.). Great Clips.alteryx.com/sites/default/files/resources/files/case-study-great-chips.pdf (accessed August 2016).
- Argyris, C. (1971). Management information systems: The challenge to rationality and emotionality. Management Science, 17(6), B-275.
- Autor, D. H. (2016). The shifts—great and small—in workplace automation. sloanreview.mit.edu/article/the-shifts-great-and-small-in-workplace-automation/(accessed August 2016).
- Blogs.ESRI.com. (2014). From customers to CxOs, Starbucks delivers world-class service. blogs.esri.com/esri/ucinsider/2014/07/29/Starbucks/(accessed August 2016).
- Blogs.microsoft.com (2015). Dartmouth-Hitchcock ushers in a new age of proactive, personalized healthcare using Cortana Analytics Suite. microsoft.com/transform/2015/07/13/dartmouth-hitchcock-ushers-in-a-new-age-of-proactive-personalized-healthcare-using-cortana-analytics-suite/(accessed August 2016).
- Blogs.pb.com. (2015). The industrial Internet, Pitney Bowes and GE.blogs.pb.com/corporate/2015/07/14/the- industrial-internet-pitney-bowes-and-ge/(accessed August 2016).
- Blogs.technet.microsoft.com (2015). ML predicts school dropout risk & boosts graduation rates. blogs.technet.microsoft.com/machinelearning/2015/06/04/ml-predicts-school-dropout-risk-boosts-graduation-rates/(accessed August 2016).
- Brynjolfsson, E., & McAfee, A. (2016). The second machine age. secondmachineage.com (accessed September 2016).
- Chae, B., Paradice, D. B., Courtney, J. F., & Cagle, C. F. (2005). Incorporating an ethical perspective into problem formulation. Decision Support Systems, 40(2), 197-212.
- CIO.in. (2015). Mankind Pharma finds an antidote in IBM solution for improving app implementation time. cio.in/solution-center/emc/55281 (accessed August 2016).
- Collins, J. (2005, June). Sysco gets fresh with RFID. RFID Journal. rfidjournal.com/articles/view?1652 (accessed September 2016).
- ComputerWeekly.com. (2016). Big-data and open source cloud technology help Gulf Air pin down customer sentiment. computerweekly.com/news/450297404/Big-data-and-open-source-cloud-technology-help-Gulf-Air- pin-down-customer-sentiment (accessed August 2016).
- Customers.microsoft.com (2015). Rockwell Automation: Fueling the oil and gas industry with IoT customers. microsoft.com/Pages/CustomerStory.aspx?recid=19922 (accessed August 2016).
- Customers.microsoft.com (2015). Tacoma Public Schools: Predicting student dropout risks, increasing graduation rates with cloud analytics. customers.microsoft.com/Pages/CustomerStory.aspx?recid=20703 (accessed August 2016).
- CxOtoday.com (2014). Cloud platform to help Pharma Co accelerate growth. cxotoday.com/story/mankind-pharmato-drive-growth-with-softlayers-cloud-platform (accessed August 2016).

- Davenport, T. H., & Harris, J. G. (2005). Automated decision making comes of age. MIT Sloan Management Review, 46(4), 83.
- Davenport, T. H., & Patil, D. J. (2012, October). Data scientist. Harvard Business Review, 70-76.
- Delen, D., Hardgrave, B., & Sharda, R. (2007). RFID for better supply-chain management through enhanced information visibility. Production and Operations Management, 16(5) 613-624.
- Delen, D., Hardgrave, B. C., & Sharda, R. (2011, April). The promise of RFID-based sensors in the perishables supply chain. IEEE Wireless Communications, 1-8.
- Demirkan, H., & Delen, D. (2013, April). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and Big Data in cloud. Decision Support Systems, 55(1), 412-421. dx.doi.org/10.1016/j.dss.2012.05.048 (accessed September 2016).
- Digit.HBS.org (2015). Starbucks :Brewing up a data storm! digit.hbs.org/submission/ Starbucks -brewing-up-a-data-storm/(accessed August 2016).
- Ekster, G. (2015). Driving investment performance with alternative data. integrity-research.com/wp-content/ uploads/2015/11/Driving-Investment-Performance-With-Alternative-Data.pdf (accessed September 2016).
- Electronic Privacy Information Center (2005). USA PATRIOT Act. epic.org/privacy/terrorism/ usapatriot (accessed September 2016).
- Elson, R. J., & LeClerc, R. (2005). Security and privacy concerns in the data warehouse environment. Business Intelligence Journal, 10(3), 51.
- Enterprise.microsoft.com. (2015). How Dartmouth-Hitchcock is challenging healthcare's status quo with Cortana Analytics. enterprise.microsoft.com/en-us/industries/health/how-dartmouth-hitchcock-is-challenging-healthcares-status-quo-with-cortana-analytics/(accessed August 2016).
- Gartner.com (2016) Gartner Says Worldwide Public Cloud Services Market Is Forecast to Reach $204 billion in 2016. http://www.gartner.com/newsroom/id/3188817 (Accessed November 2016).
- GE.com (2016). Making machines intelligent is smart business.ge.com/digital/sites/default/files/ge_ digital_predix_ pb_brochure.pdf (accessed August 2016).
- GEreports.com (2016). The power of Predix: An inside look at how Pitney Bowes is using the industrial Internet platform.gereports.com/the-power-of-predix-an-inside-look-at-how-pitney-bowes-has-been-using-the-industrial-internet-platform (accessed August 2016).
- Gnau, S. (2010). Find your edge. Teradata Magazine Special Edition Location Intelligence. teradata.com/articles/ Teradata-Magazine-Special-Edition-Location-Intelligence-AR6270/?type=ART (accessed September 2016).
- Goel, V. (2014, June 29). Facebook tinkers with users' emotions in news feed experiment, stirring outcry. New York Times. nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html (accessed September 2016).

- Hardgrave, B. C., Langford, S., Waller, M., & Miller, R. (2008). Measuring the impact of RFID on out of stocks at Wal-Mart. MIS Quarterly Executive, 7(4), 181-192.

- IBMBigdatahub.com (2015). SilverHook Powerboats: Tracking fast-moving boats in real time. http://www.ibm-Bigdatahub.com/blog/silverhook-powerboats-tracking-fast-moving-powerboats-real-time (accessed August 2016).

- IBM.com (2015). Case study: SilverHook Powerboats develops a racing app 40 percent faster. ibm.com/cloud-computing/case-studies/silverhook-powerboats.html (accessed August 2016).

- IBM.com (2014). Softlayer hosting platform reduces application implementation time by 98 percent. www-03.ibm.com/software/businesscasestudies/us/en/corp?synkey=Y979749I50926G25 (accessed August 2016).

- IBM.com (2015). Smarter care at MD Anderson. www-03.ibm.com/software/businesscasestudies/us/en/corp? synkey=H447240O66679Z38 (accessed August 2016).

- Juniper Research (2016). 'Internet of Things' connected devices to almost triple to over 38 billion units by 2020. http://www.juniperresearch.com/press/press-releases/iot-connected-devices-to-triple-to-38-bn-by-2020 (accessed September 2016).

- Kalakota, R. (2011). Analytics-as-a-service: Understanding how Amazon.com is changing the rules. practicalanalytics. wordpress.com/2011/08/13/analytics-as-a-service-understanding-how-amazon-com-is-changing-the-rules (accessed September 2016).

- Katz, J. (2006, February). Reaching for ROI on RFID. Industry Week. industryweek.com/companies-amp-executives/reaching-roi-rfid (accessed September 2016).

- Krivda, C. D. (2010). Pinpoint opportunity. Teradata Magazine Special Edition Location Intelligence. teradata.com/articles/Teradata-Magazine-Special-Edition-Location-Intelligence-AR6270/?type=ART (accessed September 2016).

- Mankindpharma.com. (n.d.). Overview. mankindpharma.com/company/companyoverview (accessed August 2016).

- Mason, R. O., Mason, F. M., & Culnan, M. J. (1995). Ethics of information management. Thousand Oaks, CA: Sage.

- McKendrick, J. (2015). SilverHook Powerboats: Tracking fast-moving boats in real time. ibmBigdatahub.com/blog/silverhook-powerboats-tracking-fast-moving-power-boats-real-time (accessed August 2016).

- McKinsey.com (2015). An executive's guide to the Internet of Things. mckinsey.com/business-functions/business-technology/our-insights/an-executives-guide-to-the-internet-of-things (accessed September 2016).

- MDanderson.org (2013). MD Anderson taps IBM Watson to power 'Moon Shots' mission. mdanderson.org/news-room/2013/10/md-anderson-ibm-watson-work-together-to-fight-cancer.html (accessed August 2016).

- Microsoft.com (n.d.). Customer stories | Rockwell Automation. microsoft.com/en-us/cloud-platform/customer-stories-rock-well-automation (accessed July 2016).

- Mintzberg, H., Lampel, J. B., Quinn, J. B., & Ghoshel, S. (2002). The strategy process, 4th ed. Upper Saddle River, NJ: Prentice Hall.

- Mobilemarketer.com (2013). Quiznos sees 20pc boost in coupon redemption via location-based mobile ad campaign. mobilemarketer.com/cms/news/advertising/14738.html (accessed September 2016).

- Moradpour, S., & Bhuptani, M. (2005). RFID field guide: Deploying radio frequency identification systems. New York: Sun Microsystems Press.

- News.pb.com (2015). GE and Pitney Bowes join forces to bring the power of the industrial Internet to the world of commerce.news.pb.com/article_display.cfm?article_id=5634 (accessed August 2016).

- O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, Crown, 2016 New York.

- Perez-Cascante, L. P., Plaisent, M., Maguiraga, L., & Bernard, P. (2002). The impact of expert decision support systems on the performance of new employees. Information Resources Management Journal 15(4), 64-78.

- Peslak, A. R. (2005). Internet privacy policies: A review and survey of the Fortune 50. Information Resources Management Journal, 18(1), 29-41.

- Ransbotham, S. (2016). How will cognitive technologies affect your organization? sloanreview.mit.edu/article/how-will-cognitive-technologies-affect-your-organization/(accessed September 2016).

- RedHat.com (2016). Gulf Air builds private cloud for Big data innovation with Red Hat Technologies. redhat.com/en/about/press-releases/gulf-air-builds-private-cloud-Big-data-innovation-red-hat-technologies (accessed August 2016).

- RedHat.com (2016). Gulf Air's Big Data innovation delivers deeper customer insight. redhat.com/en/success-stories (accessed August 2016).

- Sahin, E., Babaï, M. A., Dallery, Y., & Vaillant, R. (2007). Ensuring supply chain safety through time temperature integrators. The International Journal of Logistics Management, 18(1), 102-124.

- Snowflake.net (n.d.). Chime delivers personalized customer experience using Chime. snowflake.net/product (accessed August 2016).

- Standage, T. (2016). Special report by The Economist: The return of the machinery question. bit.ly/28X8cfD, economist.com/news/special-report/21700761-after-many-false-starts-artificial-intelligence-has-taken-will-it-cause-mass (accessed September 2016).

- Stein, J. (2011). Data mining: How companies now know everything about you. Time. time.com/time/magazine/article/0,9171,2058205,00.html (accessed September 2016).

- Swedberg, C. (2006a, October). Samworth keeps cool with RFID. RFID Journal. rfidjournal.com/article/articleview/2733/(accessed September 2016).

- Swedberg, C. (2006b, December). Starbucks keeps fresh with RFID. RFID Journal. rfidjournal.com/article/articleview/2890/(accessed September 2016).

- TheGuardian.com (2013). John Snow's data journalism: The cholera map that changed the world. theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map (accessed August 2016).
- Turck, M. (2016). Internet of Things: Are we there yet? (The 2016 IoT landscape). mattturck.com/2016/03/28/2016-iot-landscape/(accessed August 2016).
- WallStreetJournal.com (2016). What they know.wsj.com/public/page/what-they-know-digital-privacy.html (accessed September 2016).
- Welch, D. (2016, July 18-24). The battle for smart car data. Bloomberg Business Week. bloomberg.com/news/articles/2016-07-12/your-car-s-been-studying-you-closely-and-everyone-wants-the-data (accessed September 2016).
- Wheeler, C. (2014). Going Big with GIS. esri.com/esri-news/arcwatch/0814/going-Big-with-gis (accessed August 2016).
- Wikipedia.org (n.d.). Cloud computing. en.wikipedia.org/wiki/cloud_computing (accessed August 2016).
- Wikipedia.org (n.d.). Mankind Pharma. wikipedia.org/wiki/Mankind_Pharma (accessed August 2016).
- Wikipedia.org (n.d.). University of Texas MD Anderson Cancer Center. wikipedia.org/wiki/University_of_Texas_ MD_Anderson_Cancer_Center (accessed August 2016).
- Winter, R. (2008). E-Bayturnstoanalyticsasaservice.information- week.com/news/software/info_management/210800736 (accessed September 2016).
- YouTube.com (2014). Smarter care at MD Anderson. youtube.com/watch?v=savJ8VQ0kcA (accessed August 2016).
- YouTube.com. (2015). Dartmouth-Hitchcock revolutionizes the U.S. healthcare system. youtube.com/watch?v=-wVeHZNn8aU (accessed August 2016).
- YouTube.com (2016). The saving power of data. youtube.com/watch?v=rfAoKs8XxzY (accessed August 2016).

المترجم في سطور

د. محمد بن عايض بن ظافر القرني.

المؤهل العلمي:

دكتــوراه في علم المعلومات، تقنية المعلومات، جامعة شــمال تكســاس، تكســاس، الولايات المتحدة الأمريكية.

العمل الحالي:

مدير عام مركز المعلومات والوثائق، بمعهد الإدارة العامة، الرياض، المملكة العربية السعودية.

أبرز الخبرات العلمية:

- Alshare, K.A., Moqbel, M. and Algarni, M.A. (2019) The impact of trust, security, and privacy on individual's use of the internet for online shopping and social media: a multi-cultural study, Int. J. Mobile Communications,17(5), (513–536).
- Algarni, M., Khairi, A. (2017) Open Access and Digital Repositories (Arabic and African Repositories) Comparative Study. Journal of the Saudi Library and Information association, 1, 489-510.
- Algarni, M., Musa, A., & Alshare, K. (2015). Readiness to reveal personal information on social media: A Cross-cultural analysis of Nigeria, United States, Saudi Arabia, and Qatar, in proceedings of the DSI, Houston, Texas.
- Alshare, K., Algarni, M., Musa, A. (2015). Readiness to reveal personal information on social media: A Cross-cultural analysis of Nigeria, United States, Saudi Arabia, and Qatar, in proceedings of the Southwest Decision Sciences Institute, Houston, TX.
- Algarni, M., Baron, B.K., Hooks, D., Wada, A., & Wade, R.L. (2013). Using social network analysis (SNA) to enhance learning: A community engagement approach. iConference 2013 Proceedings of UNT.
- Algarni, M. (2010). Look! Libraries without Public Access Catalogue: The potentialities of GPS + RFID technologies in Academic Libraries, CULS conference, Emporia, KS.

- Dow, M., Algarni, M., Summey, T., Valenti, S., Ibrahim, A., Polepeddi, P., & Hallett, D. (2012). Infoliteracy@adistance: Creating Opportunities to Reach (Instruct) Distance Students, Journal of Library & Information Services in Distance Learning, 6(3-4), Memphis, Tennessee.
- Algarni, M., Musa, A., & Alshare, K. (2010). A Conceptual Framework For Factors That Influence Information Utilization By Project Teams: A Human Information Behavior Approach, in proceedings of the DSI, San Diego, California.

أبرز الخبرات العملية:

- **٢٠١٧م - حتى الآن:** مدير عام مركز المعلومات والوثائق، معهد الإدارة العامة، الرياض، المملكة العربية السعودية.
- **٢٠١٨م - حتى الآن:** رئيس فريق بناء وتصميم نظام المتابعة الإلكتروني، معهد الإدارة العامة، الرياض، المملكة العربية السعودية.
- **٢٠١٩م - حتى الآن:** عضو اللجنة الدائمة الرئيسية لبرنامج التحول الوطني (التدريب عن بُعد، إثرائي)، معهد الإدارة العامة، الرياض، المملكة العربية السعودية.
- **٢٠١٩م - حتى الآن:** رئيس اللجنة الدائمة للمتابعة والتوثيق لبرنامج التحول الوطني (التدريب عن بُعد، إثرائي)، معهد الإدارة العامة، الرياض، المملكة العربية السعودية.
- **٢٠١٩م - حتى الآن:** رئيس الفريق المعلوماتي لبناء وتصميم نظام ذكاء الأعمال والتحليلات ولوحات المعلوماتية الذكية، بمركز المعلومات، معهد الإدارة العامة، الرياض، المملكة العربية السعودية.
- **٢٠١٩م:** تصميم وإعداد برنامج تدريبي موجه للتنفيذيين (ذكاء الأعمال ودعم القرار)، معهد الإدارة العامة، الرياض، المملكة العربية السعودية.
- **٢٠١٩م:** تصميم وإعداد حلقة تطبيقية للقيادات (ذكاء الأعمال وصناعة القرار)، معهد الإدارة العامة، الرياض، المملكة العربية السعودية.
- **٢٠١٦ - ٢٠١٩م:** رئيس فريق الخدمات الاستشارية لدراسات اللجنة الوزارية للتنظيم الإداري (الملغاة)، معهد الإدارة العامة، الرياض، المملكة العربية السعودية.
- **٢٠١٦ - ٢٠١٧م:** مدير عام مركز التوثيق الإداري، معهد الإدارة العامة، الرياض، المملكة العربية السعودية.
- **٢٠١٣ - ٢٠١٤م:** رئيس قسم المعلومات وكبير الاستشاريين، الشركة الأمريكية للصحة العضوية - دوام كامل، دالاس، الولايات المتحدة الأمريكية.

مراجع الترجمة في سطور

د. حمود بن ظافر الدوسري.

المؤهل العلمي:

- دكتوراه في علوم الحاسب من جامعة كاردف بالمملكة المتحدة، أكتوبر ٢٠١١م.

العمل الحالي:

- أستاذ مشارك، قسم نظم المعلومات بكلية علوم الحاسب والمعلومات، جامعة الملك سعود.

أبرز الخبرات العلمية:

- Hessah Alsaif, Hmood Al-Dossari. Detecting and Classifying Crimes from Arabic Twitter Posts using Text Mining Techniques. In the International Journal of Advanced Computer Science and Applications (IJACSA) - Volume 9 No 10, October 2018.
- Shamsul Huda, Kevin Liu, Mohamed Abdelrazek, Amani Ibrahim, Sultan Alyahya, Hmood Al-Dossari, and Shafiq Ahmad. An ensemble oversampling model for class imbalance problem in software defect prediction. In IEEE Access, Volume 6, March 2018 (ISI-indexed Journal)
- Shamsul Huda, Sultan Alyahya, Md Mohsin Ali, Shafiq Ahmad, Jemal Abawajy, Hmood Al-Dossari, and John Yearwood. A Framework for Software Defect Prediction and Metric Selection. In IEEE Access, Volume 6, December 2017 (ISI-indexed Journal) 5. Hailah Al-Ballaa, Hmood Al-Dossari, Abdulrahman Mirza. Research area classification using Wikipedia and information retrieval. Proceedings of the 1st International Conference on Internet of Things and Machine Learning, ACM Page 49 – 51, October 2017
- Mohammad Mehedi Hassan, Hanouf Saad Albakr, Hmood Al-Dossari, and Amr Mohamed. Resource Provisioning for Cloud-Assisted Body Area Network in a Smart Home Environment. In IEEE Access, Volume 5, July 2017 (ISI-indexed Journal).

- H. Al-Dossari, J. Shao, A. Preece. QoS Assessment over Multiple Attributes. In IEEE 10th International Conference on Computer and Information Technology (CIT2010), pp 1456-1461. IEEE, 2010. 37. H. Al-Dossari, A. Preece, J. Shao. Improving QoS Assessment Involving Multiple Attributes. In 26th British National Conference on Databases PhD Forum (BNCOD 2009). Birmingham, 2009.

أبرز الخبرات العملية:

- **٢٠١٧م - حتى الآن:** أســتاذ مشارك في قسم نظم المعلومات بكلية علوم الحاسب والمعلومات في جامعة الملك سعود.
- **٢٠١٤ - ٢٠١٨م:** رئيس قسم نظم المعلومات بكلية علوم الحاسب والمعلومات في جامعة الملك سعود.
- **٢٠١٣م:** أستاذ مساعد في قسم نظم المعلومات بكلية علوم الحاسب والمعلومات في جامعة الملك سعود.
- **٢٠١٢ - ٢٠١٣م:** مدير شــعبة الحاســب الآلي في قســم مهارات تطوير الذات بعمادة الســنة التحضيرية في جامعة الملك سعود.
- **٢٠١١ - ٢٠١٢م:** رئيس قسم اللغة الإنجليزية بكلية المعلمين بجامعة الملك سعود.
- **٢٠١١ - ٢٠١٢م:** أستاذ مساعد بقسم الحاسب الآلي بكلية المعلمين في جامعة الملك سعود.
- **٢٠٠٦ - ٢٠٠٧م:** مدير المدرسة السعودية بكادرف في المملكة المتحدة.
- **٢٠٠٥ - ٢٠١١م:** محاضر بقسم الحاسب الآلي بكلية المعلمين في جامعة الملك سعود.
- **٢٠٠١ - ٢٠٠٥م:** معيد بقسم الحاسب الآلي بكلية المعلمين، وزارة التربية والتعليم.
- **١٩٩٩ - ٢٠٠١م:** معيد بكلية الملك فيصل الجوية، وزارة الدفاع والطيران.
- **١٩٩٨ - ١٩٩٩م:** مبرمج على المرتبة السابعة في وكالة الضمان الاجتماعي بوزارة العمل والشؤون الاجتماعية.
- عضو البرنامج الوطني للتحول في التعاملات الإلكترونية الحكومية يسر.
- مستشار غير متفرغ في الهيئة العليا لتطوير مدينة الرياض.
- رئيس لجنة إدارة الجودة بقسم الحاسب الآلي بكلية المعلمين في جامعة الملك سعود.

تم التصميم والإخراج الفني والطباعة في

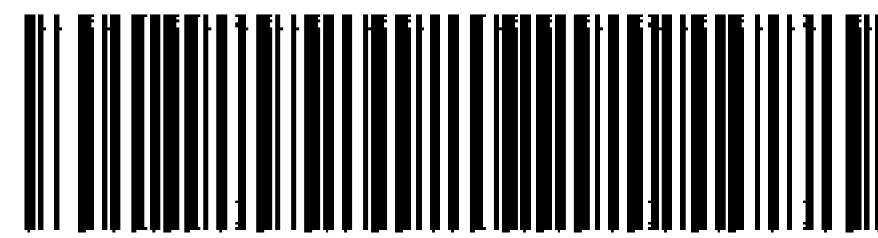
الإدارة العامة للطباعة والنشر - معهد الإدارة العامة ١٤٤١هـ

هذا الكتاب:

تطوَّر مصطلح التحليلات إلى مصطلحات حديثة أخرى، مثل علم البيانات، وهو أحدث تجسيدٍ للتعلُّم العميق وإنترنت الأشياء؛ إذ تقوم أغلب المنظمات - سواء خاصة أو عامة - بإنشاء وحدات تنظيمية جديدة تركِّز على التحليلات التي تدعم صُناع القرار معلوماتيًّا وتُسهِم في رفع كفاءة وفعالية عملياتها؛ وذلك بالاعتماد على أحدث الأدوات المعلوماتية الذكية، مثل ذكاء الأعمال، والذي يستخلص المعلومات بطريقةٍ مبتكرة من تيارات البيانات المتنوعة.

يُقدِّم هذا الكتاب منظوراً إداريًّا لتحليلات الأعمال في ثمانية فصول، تبدأ بنظرة عامة على ذكاء الأعمال، والتحليلات، وعلم البيانات، ثم ينتقل إلى التحليلات الوصفية وتتضمَّن طبيعة البيانات، والنمذجة الإحصائية، وتصوير البيانات، وذكاء الأعمال، ومستودعات البيانات ومصطلحات حديثة، مثل: بحيرات البيانات، متبوعة بالتحليلات التنبؤية وتتضمَّن عملية وطرق وخوارزميات التنقيب في البيانات، وتحليلات النص والويب ووسائل التواصل الاجتماعي، ويليها التحليلات التوجيهية، مثل الأمثلية والمحاكاة، ثم استعراض المفاهيم والأدوات الأساسية للبيانات الضخمة، وينتهي الكتاب بالتعرُّف على الاتجاهات المستقبلية والخصوصية والاعتبارات الإدارية في التحليلات.

ختاماً؛ فإنَّ هذا الكتاب يُعَدُّ مرجعاً أساسيًّا للأكاديميين والمهنيين على كافة مستوياتهم وتخصصاتهم؛ مما جعله من الكتب الرائدة في السوق. وقد تميَّزت هذه الطبعة المنقحة من هذا الكتاب بالشمولية والوضوح وسهولة الاستخدام وتقديم مواد دقيقة ومُحدَّثة غير متوفرة في أيِّ نصٍّ آخر؛ إضافةً للشرح التفصيلي المُعزَّز بحالات وأمثلة من عالم الواقع، مثل أفضل الممارسات لتوظيف الأفراد الأذكياء للأدوات المعلوماتية لتحسين الأداء والعلاقات في منظمات الأعمال والحكومات والمنظمات غير الرِّبحية؛ مما ارتقى بالكتاب إلى مستوى الدَّعم المعلوماتي للعمل الإداري؛ وبشكلٍ أساسي صُنع القرار.



9786038276235

تصميم وإخراج وطباعة الإدارة العامة للطباعة والنشر - معهد الإدارة العامة ١٤٤١هـ